# Quantization Techniques in Linearly Precoded Multiuser MIMO system with limited feedback

by

Muhammad Nazmul Islam

A thesis submitted in conformity with the requirements
for the degree of Masters of Applied Science
Graduate Department of Electrical & Computer Engineering
University of Toronto

# Abstract

Quantization Techniques in Linearly Precoded Multiuser MIMO system with limited feedback

Muhammad Nazmul Islam

Masters of Applied Science

Graduate Department of Electrical & Computer Engineering

University of Toronto

2010

Multi-user wireless systems with multiple antennas can drastically increase the capacity while maintaining the quality of service requirements. The best performance of these systems is obtained at the presence of instantaneous channel knowledge. Since uplink-downlink channel reciprocity does not hold in frequency division duplex and broadband time division duplex systems, efficient channel quantization becomes important. This thesis focuses on different quantization techniques in a linearly precoded multi-user wireless system.

Our work provides three major contributions. First, we come up with an end-to-end transceiver design, incorporating precoder, receive combining and feedback policy, that works well at low feedback overhead. Second, we provide optimal bit allocation across the gain and shape of a complex vector to reduce the quantization error and investigate its effect in the multiuser wireless system. Third, we design an adaptive differential quantizer that reduces feedback overhead by utilizing temporal correlation of the channels in a time varying scenario.

# Acknowledgements

First and foremost, I would like to thank Professor Raviraj Adve, for his guidance and supervision throughout the thesis. He introduced me to the thesis topic, helped me in grasping the materials and guided me throughout this journey. His encouragement, support and insight are greatly appreciated.

I would like to thank Professor T. J. Lim for providing valuable suggestions during my thesis proposal. Here I would like to mention a number of senior students who guided me throughout various phases of the thesis. Adam Tenenbaum introduced me to the concepts of linear precoding and almost acted as my "virtual co-supervisor" during early parts of my thesis. KV Srinivas was kind enough to allow me to discuss new ideas with him. My proofs on optimal bit allocation could have never been completed without the help of Behrouz Khonevis. I am really grateful to all of you.

I am also thankful to members of the communication group for their friendship and support over last two years. The monotonous hours at BA 7114 became enjoyable due to the presence of Gokul Sridharan, Ehsan Karamad, Wael Louis, Hassan Masoom, Ali Khanafer, Kianoush Hosseini, Colin, William Chou, Sachin Kadloor, Amir Aghaei and Siddarth Hari. Whereever I move from here, I will always remember and miss BA 7114.

I specially want to thank my parents for being the beacon of my life. You have been helping me at every stage of my life. It is only because of your tremendous sacrifice that I have come up to this point in my career. I also appreciate the support that I have received from my two sisters and their families over the years. I am really grateful to Nafisa, my wife, for her love and support over the last two years. Toronto, a wonderful city itself, became more wonderful due to your presence.

Funding for this research was provided by LG Electronics, Korea.

# Contents

# List of Tables

# List of Figures

# Notation

| | |
|---|---|
| $\lvert \cdot \rvert$ | Magnitude of a complex number |
| $\lVert \cdot \rVert$ | Euclidean norm of a vector |
| $\lVert \cdot \rVert_F$ | Frobenius norm of a matrix |
| $(\cdot)^T$ | Transpose |
| $(\cdot)^H$ | Conjugate transpose (Hermitian) |
| $x$ | Scalar |
| $\mathbf{x}$ | Vector |
| $\mathbf{X}$ | Matrix |
| $[x_1, x_2, \cdots, x_N]$ | Concatenation of scalars into row vector |
| $[\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]$ | Concatenation of column vectors into matrix |
| $[\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_N]$ | Concatenation of matrices |
| $\text{diag}\,[\cdot]$ | Concatenation of entries into (block) diagonal matrix |
| $\mathcal{CN}(\mu, \mathbf{\Sigma})$ | Multivariate normal distribution with mean $\mu$ and covariance $\mathbf{\Sigma}$ |
| $tr[\cdot]$ | Trace of a matrix |
| $E[\cdot]$ | Statistical expectaion |
| $\mathbf{A}(:, L_1)$ | Concatenation of first $L_1$ columns of matrix $\mathbf{A}$ |
| $\angle(\mathbf{b}, \mathbf{C})$ | Angle between $\mathbf{b}$ and the subspace spanned by the columns of $\mathbf{C}$ |
| $SA(A)$ | Surface area of $A$ |
| $[1, K]$ | A set that contains $(1, 2, 3, \cdots, K-1, K)$ |

# Chapter 1

# Introduction

## 1.1 Motivation

Wireless communication systems face increasing demands in terms of quality of service and number of users. The implementation of multiple input multiple output (MIMO) systems can play a significant role in meeting this demand. MIMO systems can be used in increasing system reliability as well as providing increased data rates. The availability of channel state information (CSI) at the transmitter increases the capacity of MIMO systems. However, in frequency division duplex (FDD) and broadband time division duplex (TDD) systems, the CSI must be estimated at the receiver, quantized and fed back to the transmitter. Clearly, there is a tradeoff between the feedback rate and accuracy of the CSI. This thesis investigates different aspects of quantization of CSI, especially in the context of multiuser MIMO communication.

The field of wireless communication has gone through rapid development in the last two decades. Two major concerns in designing the wireless systems are the reliability of the transmission and the capacity of the network. The number of users that can fit in a particular transmission is limited by the bandwidth and time resources. Current technologies try to serve multiple users by spreading them over time and frequency in

code division multiple access (CDMA) systems or by distributing them over different time slots in time division multiple access (TDMA) systems.

MIMO systems allow for spatial dimensions in the network enabling for space division multiple access (SDMA). SDMA enables the simultaneous access to different users in the same time and frequency slot using independent paths.

The capacity of the MIMO systems increase linearly with the minimum number of the transmit and receive antennas. In next generation wireless networks, users are expected to communicate different streams such as text, audio & video simultaneously with each other. These data streams have different quality of service (QoS) requirements. By creating several independent data paths between the transmitter and the receiver and using suitable coding techniques, MIMO systems can meet these demands.

Fading is one of the most important challenges that need to be addressed while working in the wireless environment. Fading refers to the random fluctuations of the wireless environment and it is mostly attributed to the scatterers located between the base station and the user. The mobility of the user changes the pattern of scattering of the received data and thus causes fluctuation in the power of the received signal. MIMO systems play a key role in mitigating the ill-effects of fading.

With CSI at the transmitter, MIMO systems also allow a single transmitter to communicate with multiple receivers on the same time/frequency channel. However, the broadcast nature of wireless communication leads to multiuser interference. Precoding at the transmitter based on the CSI helps to combat both fading and multiuser interference. Precoding uses the spatial dimension of the channel to combat fading and MUI. Other coding techniques like space-time coding use the temporal dimension to combat fading and do not require the channel state information.

In our work, we only focus on linear precoding i.e. our algorithm will only consist of matrix multiplications and additions. This reduces the complexity of data transmission.

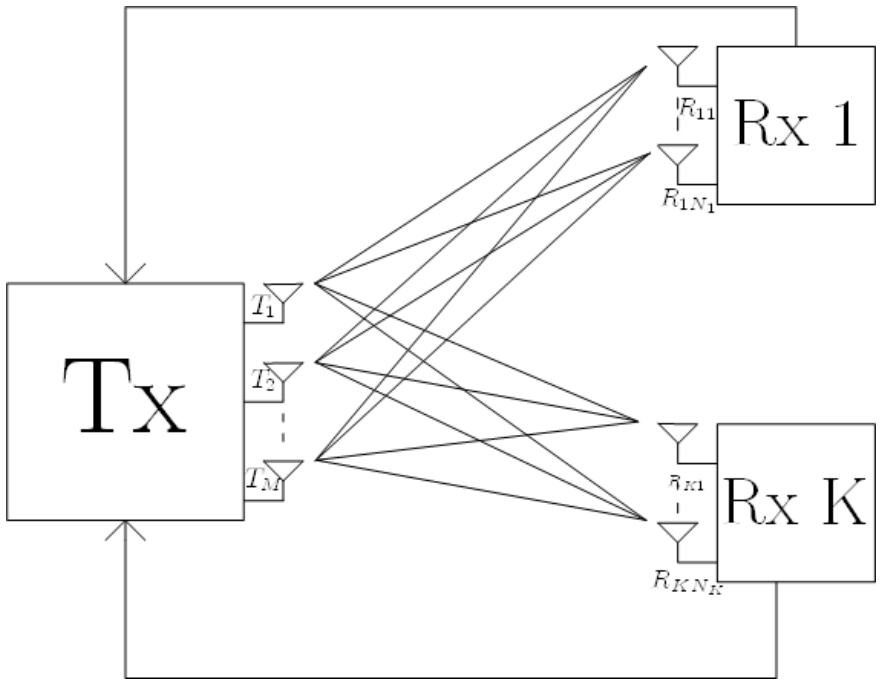



Figure 1.1: Generic model of multiuser MIMO channel with limited feedback

## 1.2 Key Challenges

We focus on a linear precoding based multiuser MIMO system that has quantized channel knowledge at the transmitter. The transmitter is a base station (BS) communicating with multiple users. Our system model is shown in Fig. 1.1. Here, $M$, $K$ & $N_i$ represent the total number of transmit antennas, the total number of users and the number of receive antennas of the $i^{\text{th}}$ user respectively. Each user has several antennas and may receive more than one data stream. Each data stream is assumed independent of the others and users do not co-operate in the decoding of the data. Each user estimates its own channel and sends back the CSI via the quantized feedback path.

Feedback of the CSI to the transmitter is an overhead to the communication and as such this overhead should be minimized. There is a trade-off between the ability to suppress multiuser interference and the rate of CSI feedback. The central aim of this thesis is to develop practical quantization schemes that perform well at low feedback

rates. In this regard, we make three specific contributions:

1. To develop practical quantization schemes that perform better at low rate feedback.

2. To reduce feedback overhead by utilizing temporal correlation of previous channel instants. This will allow the proposed schemes to be utilized in time varying channels.

3. To provide optimal bit allocation across gain and shape of the effective vector downlink channel in a MIMO system.

## 1.3 Thesis Structure

This thesis is organized as follows. Chapter 2 provides a detailed survey of previous works on channel quantization and receive combining techniques, optimal bit allocation across gain and shape of vector and adaptive differential quantization in time varying channels. Chapter 3 presents our work on MIMO receive combining. We analyze the quantization error and overall sum mean squared error (SMSE) and show how the proposed algorithm compares with the existing limited feedback based quantization techniques. Chapter 4 finds the quantization error variance of gain and shape of a complex vector for a given number of allocated bits. Thereafter, we provide optimal bit allocation across the gain and shape of the vector. Chapter 5 develops the proposed adaptive differential quantization techniques and shows how our algorithm can lead to the reduction of several kBit/sec feedback overhead in a time varying scenario. Finally, Chapter 6 summarizes contributions of the thesis and suggests possibilities of future work.

## 1.4 Publications

This thesis has led to the following publications:

1. M. N. Islam and R.S. Adve, SMSE precoder design in a multiuser MISO system with limited feedback, 2010 Queens Biennial Symposium [26].

2. M. N. Islam and R.S. Adve, Linear transceiver design in a multiuser MIMO system with quantized channel state information, 2010 IEEE ICASSP [25].

3. M. N. Islam and R.S. Adve, Tranceiver design using linear precoding in a multiuser system with limited feedback, IET Transactions on Communications [27].

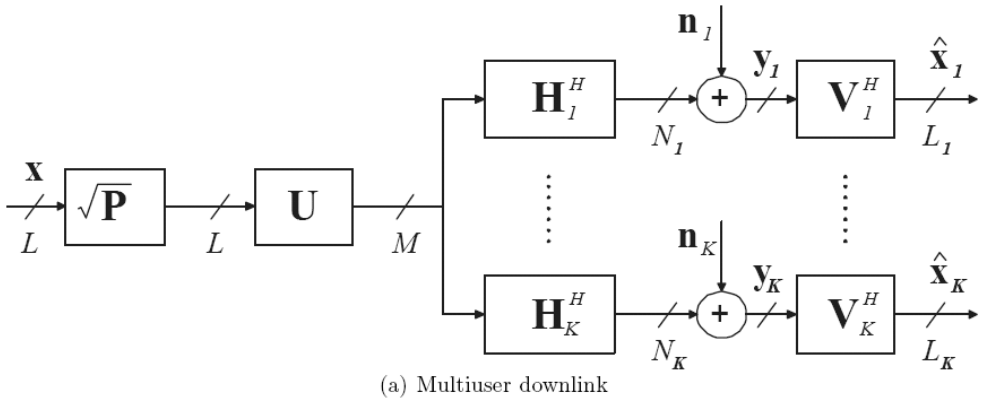
# Chapter 2

# Background

The advantages of spatial diversity and multiplexing have led to the investigation of multi user (MU) multiple input single output (MISO) and multiple input multiple output (MIMO) wireless communication systems. Precoding allows to combat fading and retain the advantages of MISO and MIMO systems. Linear precoding is attractive due to its linear nature. The best performance of linear precoding can be achieved when channel state information (CSI) is available at the transmitter. This literature review covers three parts. First, we review on different quantization techniques in a linearly precoded multiuser MIMO system. Second, we focus on optimal bit allocation across gain and shape feedback in a multiuser channel. Finally, we review previous works that have dealt with feedback overhead reduction in a time varying channel.

## 2.1 Quantization Techniques in Multiuser MIMO Systems

In our work, we assume that a single transmitter (base station) is communicating with several receivers (mobile stations) of a broadcast channel. Each user may have a single or multiple antennas. We focus on the scenario where each user receives different data




(a) Multiuser downlink

Figure 2.1: Block diagram of multiuser MIMO downlink

streams. We do not assume co-ordination among different receivers. Therefore, receive antennas of a given user can only co-ordinate with each other.

### 2.1.1 Linear Precoded Systems

We start our work with a detailed description of the system model. The system model introduced in this section will be used throughout the thesis.

#### System Model

Consider a single base station equipped with $M$ transmit antennas communicating with $K$ independent users. User $k$ has $N_k$ antennas and receives $L_k$ data streams. Let $L = \sum_k L_k$, $N = \sum_k N_k$. To ensure linear independence among the data streams, we assume $L \leq M$ and $L_k \leq N_k$.

We clarify a particular use of symbol that will be used throughout this thesis. In our system model, each user can receive multiple data streams. Therefore, our overall system design is based on both per user operation and per data stream operation. Unless stated otherwise, the notation $k$ and $i$ will be used to denote a user and a data stream respectively. The $i^{\text{th}}$ data stream is processed by a unit norm linear precoding vector $\mathbf{u}_i$.

The global precoder can be formulated as follows,

$$\begin{aligned} \mathbf{U} &= [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_L] && (2.1)\\ &= \left[\mathbf{u}_{1_1}, \mathbf{u}_{1_2}, \cdots, \mathbf{u}_{1_{L_1}}, \mathbf{u}_{2_1}, \mathbf{u}_{2_2}, \cdots, \mathbf{u}_{2_{L_2}}, \cdots, \mathbf{u}_{K_1}, \cdots, \mathbf{u}_{K_{L_K}}\right] && (2.2)\\ &= [\mathbf{U}_1, \mathbf{U}_2, ..., \mathbf{U}_K] && (2.3) \end{aligned}$$

Here, $\mathbf{U}_k$ denotes the transmit filters of the $k^{\text{th}}$ user, i.e., $\mathbf{U}_k = \left[\mathbf{u}_{k_1}, \cdots, \mathbf{u}_{k_{L_k}}\right]$. Here, $i = k_1, \cdots, k_{L_k}$ denotes the data streams of the $k^{\text{th}}$ user i.e. $k_{L_k}$ represents the $\mathrm{L}_k^{th}$ stream of the $\mathrm{k}^{th}$ user.

Section 3.4.2 will show that the receive combiner, i.e., decoding filter and quantization policy is designed from users' perspective. Therefore, we use the symbol policy of (2.2) in receive combiner section. On the other hand, as will be shown in section 3.5, we design the transmit filter from data stream's perspective. Therefore, we use the symbol policy of (2.1) in the transmit precoder design.

Fig. 2.1 shows the block diagram of the proposed system in the downlink. Let $\mathbf{p} = [p_1, p_2, .., p_L]^T = [\mathbf{p}_1, \mathbf{p}_2, .., \mathbf{p}_K]^T$ be the powers allocated to the $L$ data streams and the $K$ users. Here, $\mathbf{p}_k = \left(p_{k_1}, \cdots, p_{k_{L_k}}\right)^T$. Define the downlink power matrix $\mathbf{P} = diag(\mathbf{p})$. The transmitter operates under the constraint $||\mathbf{p}||_1 \leq P_{max}$ where $P_{max}$ is the total available power.

The data vector $\mathbf{x} = [x_1, ...., x_L]^T = \left[\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_K^T\right]^T$, includes all $L$ data streams to the $K$ users. The $N_k \times M$ block fading channel, $\mathbf{H}_k^H$, between the BS and the user is assumed to be flat. The global channel matrix is $\mathbf{H}^H$, with $\mathbf{H} = [\mathbf{H}_1, ..., \mathbf{H}_k]$. User $k$ receives

$$\mathbf{y}_k^{DL} = \mathbf{H}_k^H \mathbf{U}\sqrt{\mathbf{P}}\mathbf{x} + \mathbf{n}_k, \tag{2.4}$$

where $\mathbf{n}_k$ represents the zero mean additive white Gaussian noise at the $k^{\text{th}}$ receiver with $E\left[\mathbf{n}_k\mathbf{n}_k^H\right] = \sigma^2\mathbf{I}_{N_k}$. We also assume, $E\left[\mathbf{x}\mathbf{x}^H\right] = \mathbf{I}_L$. To estimate its own transmitted

symbols, from $\mathbf{y}_k^{DL}$, user $k$ forms

$$\hat{\mathbf{x}}_k = \mathbf{V}_k^H \mathbf{y}_k^{DL} \tag{2.5}$$

$$= \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{U}\sqrt{\mathbf{P}}\mathbf{x} + \mathbf{V}_k^H \mathbf{n}_k \tag{2.6}$$

Here $\mathbf{V}_k = \left[\mathbf{v}_{k_1}, \cdots \mathbf{v}_{k_{L_k}}\right]$ is the $N_k \times L_k$ decoder matrix for user $k$. $\mathbf{v}_{k_j}$ denotes the decoding vector of the j$^{th}$ stream of the k$^{th}$ user. Let $\mathbf{V}$ be the $N \times L$ block diagonal global decoder matrix,

$$\mathbf{V} = diag\left(\mathbf{V}_1, ..., \mathbf{V}_K\right) \tag{2.7}$$

Hence, combining all users,

$$\mathbf{y} = \mathbf{H}^H \mathbf{U}\sqrt{\mathbf{P}}\mathbf{x} + \mathbf{n} \tag{2.8}$$

$$\hat{\mathbf{x}} = \mathbf{V}^H \mathbf{H}^H \mathbf{U}\sqrt{\mathbf{P}}\mathbf{x} + \mathbf{V}^H \mathbf{n} \tag{2.9}$$

where, $\mathbf{n} = \left[\mathbf{n}_1^T, \mathbf{n}_2^T, \ldots, \mathbf{n}_K^T\right]^T$.

### 2.1.2 Design Focus

The design of ($\mathbf{P}$, $\mathbf{U}$ and $\mathbf{V}$) in the above stated system model has been investigated with different criteria. These criteria include maximizing throughput, minimizing mean square error (MSE) under a total power constraint [51], or minimizing total transmitted power while satisfying qualtity of service (QoS) constraints (e.g. signal-to-interference plus noise ratio (SINR) for each stream) [6].

In this work, we focus on minimizing SMSE. In the following sub-sections, we provide a brief literature review on the linear precoder design with MMSE objective.

#### Original SMSE objective

Let $\mathbf{E}_k^{DL}$ denotes the $L_k \times L_k$ error covariance matrix of user k in the downlink, where

$$\mathbf{E}_k^{DL} = E\left[\left(\hat{\mathbf{x}}_k - \mathbf{x}_k\right)\left(\hat{\mathbf{x}}_k - \mathbf{x}_k\right)^H\right] \tag{2.10}$$

The diagonal entries of $\mathbf{E}_k^{DL}$ are the MSEs of the $L_k$ substreams of user $k$. Therefore, $SMSE_k^{\mathrm{DL}} = tr\left[\mathbf{E}_k^{DL}\right]$. The SMSE minimization problem can be formulated as,

$$min_{\mathbf{p},\mathbf{U},\mathbf{V}} \sum_{k=1}^{K} tr\left[E_k^{DL}\right] \tag{2.11}$$

subject to : $||\mathbf{p}||_1 \leq P_{max}$.

**Precoder Design with Full Channel Knowledge**

Section 2.1.1 shows that the design of $\mathbf{U}$ and $\mathbf{V}$ depend on the channel knowledge $\mathbf{H}$. Therefore, the primary works on linear precoder design in the downlink assumed perfect channel knowledge.

Tenenbaum and Adve [55] focused on this work from the downlink perspective. The authors used iterative joint optimization (IJO) and sequential quadratic programming (SQP) techniques in their design. The optimization of (2.11) with respect to $\mathbf{U}$ is a non-convex problem in the downlink. Since both IJO and SQP solve (2.11) from downlink perspective, they suffer from increased computational complexity.

After the works in [55], there have been several works on the SMSE linear precoder design that have exploited a duality between the downlink and a virtual uplink. Before reviewing those works, we provide a description of duality, in terms of SMSE.

Let us imagine a virtual uplink channel model where the users transmit data to the base station. Figure 2.2 illustrates the linear processing involved in the virtual uplink of the system. In this uplink, the transmit powers are $\mathbf{q} = [q_1, .., q_L]^T$ for the $L$ data streams, while the matrices $\mathbf{U}$ and $\mathbf{V}$ become the receive and transmit matrices respectively. The global virtual uplink power allocation matrix $\mathbf{Q}$ is defined as, $\mathbf{Q} = diag(\mathbf{q})$ where $||\mathbf{q}||_1 \leq P_{max}$. Therefore,

$$\mathbf{y}^{UL} = \sum_{j=1}^{L} \mathbf{H}_j \mathbf{V}_j \sqrt{q_j} x_j + \mathbf{n} \tag{2.12}$$

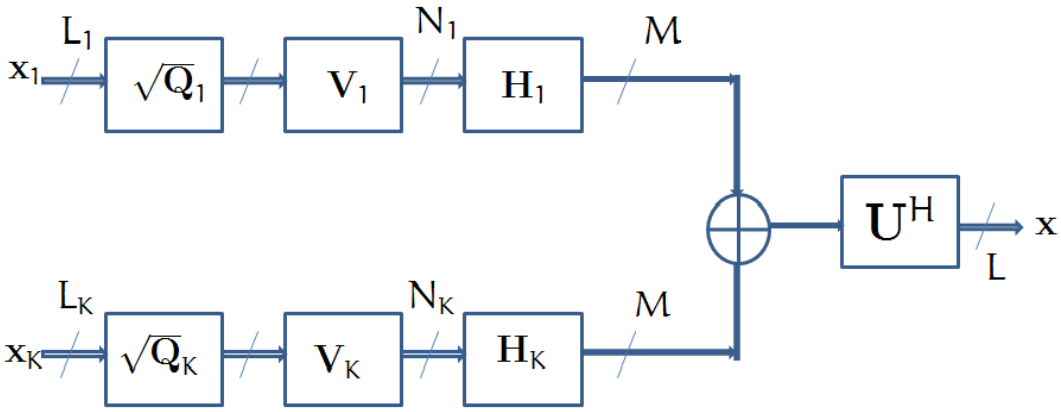



Figure 2.2: Block diagram of multiuser MIMO uplink

$$\hat{\mathbf{x}}_k^{UL} = \sum_{j=1}^{L} \mathbf{U}_k^H \mathbf{H}_j \mathbf{V}_j \sqrt{q_j} x_j + \mathbf{U}_k^H \mathbf{n} \tag{2.13}$$

Now, let $\mathbf{E}_k^{UL}$ denote the $L_k \times L_k$ error covariance matrix of user $k$ in the uplink, where

$$\mathbf{E}_k^{UL} = E\left[\left(\hat{\mathbf{x}}_k^{UL} - \mathbf{x}_k^{UL}\right)\left(\hat{\mathbf{x}}_k^{UL} - \mathbf{x}_k^{UL}\right)^H\right], \tag{2.14}$$

and $SMSE^{\mathrm{UL}} = \sum_{k=1}^{K} tr(\mathbf{E}_k)$. The SMSE minimization problem in the uplink, takes the following form,

$$\begin{aligned} min_{\mathbf{q},\mathbf{U},\mathbf{V}} \quad & \sum_{k=1}^{K} tr\left[\mathbf{E}_k^{UL}\right] \\ subject\,to: \quad & ||\mathbf{q}||_1 \leq P_{max} \end{aligned} \tag{2.15}$$

In a multiuser MISO system, the design of $\mathbf{U}$ and the allocation of $\mathbf{p}$ in (2.15) take the form of a standard Weiner filter and a convex optimization problem formulation respectively [51]. Shi and Schubert [51] prove that, for a given precoding vector and total power budget, normalized mean square error between downlink and virtual uplink can be shown to be equal. The use of uplink-downlink duality simplifies linear precoder design.

Khachan et al. [32] and Schubert et al. [49] extended the work of [51] to multiuser MIMO systems. In the MIMO case, for a given $\mathbf{U}$ and $\mathbf{V}$, there exists $\mathbf{p}$ and $\mathbf{q}$ such that

$||\mathbf{p}||_1 = ||\mathbf{q}||_1 = P_{max}$ and $MSE_i^{UL} = MSE_i^{DL}$. Schubert et. al. [49] used the following solution for the downlink precoding filter (i.e., virtual uplink decoding filter),

$$\mathbf{U} = \left(\mathbf{HVQV}^H\mathbf{H}^H + \sigma^2\mathbf{I}\right)^{-1}\mathbf{HVQ} \tag{2.16}$$

The authors find the optimum covariance matrix, $\mathbf{R} = \mathbf{VQV}^H$ using a semidefinite optimization problem. $\mathbf{V}$ and $\mathbf{Q}$ were found using an eigenvalue decomposition from $\mathbf{R}$.

On the other hand, Khachan et. al. used a rank-one minimization algorithm, by iterating between optimal $\mathbf{V}$ and $\mathbf{Q}$ to minimize the uplink SMSE in (2.14) until convergence. Based on the designed $\mathbf{V}$ and $\mathbf{Q}$, the authors prove the uplink-downlink duality in terms of SMSE and find the downlink precoding filter. The optimal $\mathbf{p}$ can then be found through a transformation. In fact, a more recent result shows that this transformation is not required and at the MMSE, $\mathbf{p} = \mathbf{q}$ [57]. We use this in our work.

All these works assumed perfect channel knowledge at the base station. Shenouda and Davidson [50] and Ding [13] investigated SMSE precoder design with channel uncertainty. They assume the following channel model in their work,

$$\mathbf{H} = \hat{\mathbf{H}} + \tilde{\mathbf{H}} \tag{2.17}$$

Here, $\hat{\mathbf{H}}$ contains the channel state information that is available at the base station. $\tilde{\mathbf{H}}$ denotes the channel uncertainty whose entries are assumed to follow a Gaussian probability distribution. The channel uncertainty can arise due to imperfect channel estimation or limited feedback. Based on this model, both [50] and [13] found the prcoding vector and power allocation matrix to minimize SMSE. We worked on a similar problem independently in our research project. The description of this part of our work is described in detail in Chapter 3.

Both Shenouda & Davidson and Ding's work assume a generalized model of channel uncertainty in the precoder design. We specifically focus on limited feedback scenario in our work. Therefore, we assume that the receivers are allocated a finite number of bits to convey the channel state information to the base station. We also assume perfect channel

estimation and delay-less noise free feedback. Therefore, in our work, the receivers have complete channel knowledge whereas the base station contains quantized channel state information. In this regard, the focus of this work is different from that in [13, 50].

In the following two subsections, we provide a brief literature review on topics related with quantization.

## 2.2 Different Types of Quantization

In the available literature, scalar quantization [9,12], vector quantization (VQ) [5,30] and matrix quantization [10,46] have all been used to quantize CSI. It is now well established in the single user, single data stream, case that projecting the MIMO channel to an appropriate vector downlink channel yields better performance than full channel scalar quantization with same feedback overhead [40]. This has led to considerable research in VQ, which reduces the feedback overhead by allocating bits in the proper vector downlink channel. In VQ, to send $B$ feedback bits as the channel index to the BS, each user needs a codebook with $2^B$ code vectors. Grassmannian line packing [41], VQ using mean squared error (MSE) as the optimality criterion [11] and random vector quantization (RVQ) [29] have been the most popular approaches in VQ.

We utilize all the feedback bits to quantize the shape of the channel in chapter 3. Mean square inner product (MSIP) based vector quantization [48] leads to a higher inner product between the original and quantized channel compared to MSE based VQ. Also, its algorithm converges faster compared to Grassmannian line packing [41]. Therefore, we use MSIP based VQ in this part of our work.

We quantize the gain and shape separately and use euclidean distance based feedback in Chapter 4. Since the optimal codebook is not known in this case, we use random VQ based on the MSE as the feedback method.

Quantization schemes require a measure of the distance between two vectors. Two

popular choices are the chordal and Euclidean distance.

### Chordal Distance

The chordal distance between any two vector $\mathbf{c}_1$ and $\mathbf{c}_2$ depends on the sine of the angle, $\theta_{1,2}$, between these two vectors [10]. This distance is expressed as,

$$d_c(\mathbf{c}_1, \mathbf{c}_2) = \sin(\theta_{1,2}) = \sqrt{1 - |\mathbf{c}_1^H \mathbf{c}_2|^2} \tag{2.18}$$

### Euclidean Distance

The Euclidean distance between any two vector $\mathbf{c}_1$ and $\mathbf{c}_2$ is expressed as,

$$d_e(\mathbf{c}_1, \mathbf{c}_2) = ||\mathbf{c}_1 - \mathbf{c}_2||_2 \tag{2.19}$$

Chordal distance ensures a higher inner product between the original and quantized channel in a limited feedback scenario [48]. As will be shown in Chapter 3, with an MMSE receiver, chordal distance based feedback outperforms its Euclidean counterpart [48]. This led us to use chordal distance based feedback in chapter 3.

On the other hand, Euclidean distance has a one-to-one mapping with SMSE precoder. We investigated the optimal bit allocation problem with a theoretical perspective in chapter 4. We therefore used Euclidean distance in the feedback policy of chapter 4.

## 2.3 Vector Quantization Schemes

### Grassmanian Line Packing

Grassmanian line packing is the problem of optimal packing of a one-dimensional subspace [41]. Let us consider the space of unit-norm channel vectors $\mathcal{H}_m$. Let us define an equivalence relation between two unit norm vectors $\mathbf{c}_1 \in \mathcal{H}_m$ and $\mathbf{c}_2 \in \mathcal{H}_m$ by $\mathbf{c}_1 \equiv \mathbf{c}_2$ if for some $\theta \in [0, 2\pi], \mathbf{c}_1 = e^{j\theta}\mathbf{c}_2$. The equivalence relation says that two vectors are

equivalent if they are on the same line in $\mathbb{C}^m$. The complex *Grassmannian manifold* is the set of all one dimensional subspaces of the space $\mathbb{C}^m$.

The distance metric in a Grassmannian manifold is the chordal distance. The Grassmanian line packing problem is the problem of finding $N$ lines in $\mathbb{C}^m$ that has maximum minimum distance between any pair of lines [41]. The minimum distance of a packing is the sine of the minimum angle between any pair of lines.

$$\delta(C) = \min_{1 \le i \le j \le N} \sqrt{1 - |\mathbf{c}_i^H \mathbf{c}_j|^2} = \sin(\theta_{\min}) \tag{2.20}$$

Here, $\theta_{min}$ is the smallest angle between any pair of lines [41]. Having solved for the optimal line packing, each vector then acts as a code vector in the quantization process.

**Random Vector Quantization**

In a random vector quantization (VQ) codebook, the code vectors are uniformly and independently distributed in $\mathbb{C}^M$. The performance is analyzed by averaging over the distribution of all possible random codebooks. The distortion can be defined both in terms of chordal or Euclidean distance.

In our work, we use two different vector quantization, namely mean-squared-inner -product (MSIP) based VQ and product VQ. In product VQ, the gain (norm) and shape (shape in complex space) of the channel are quantized separately. The quantized channel consists of the product between the quantized norm and the quantized shape. We provide the description and the advantages of those VQ techniques in the appropriate chapters.

## 2.4 Receive Combining Schemes

In a MU MISO system, users can feed back the channel vectors using VQ. However, in the MIMO case, one option is to combine the receive antennas to convert the MIMO channel to the effective vector downlink MISO channel. Since the receivers cannot cooperate, the quantization scheme of each user is independent of the other. In the recent literature on

limited feedback, a lot of works have been done on receive combining in MIMO systems. Note that, *all these receive combining schemes are implemented prior to the precoder design*. Besides, these receive combiners are designed from the perspective of increasing expected SINR. Here, we provide a brief review on receive combining in MIMO systems.

**EigenBased Combining (EBC)**

Let us assume a single-user MIMO channel where the user receives $L_1$ streams. Let $\mathbf{H}_1$, $\mathbf{U}_1$ and $\mathbf{V}_1$ denote the channel, precoding and decoding matrices respectively. Now, let the singular value decomposition of $\mathbf{H}_1$ be represented through the following equation,

$$\mathbf{H}_1 = \mathbf{A}_1 \Sigma_1 \mathbf{B}_1^H \tag{2.21}$$

Here, $\mathbf{A}_1$ and $\mathbf{B}_1$ represent the left and right singular matrix of $\mathbf{H}_1$. If provided with full CSI, in EigenBased Combining, $\mathbf{V}_1 = \mathbf{B}_1(:, 1 : L_1)$, $\mathbf{U}_1 = \mathbf{A}_1(:, 1 : L_1)$. When $L_1 = 1$, the scheme is called Maximum Eigenmode Transmission (MET) [5].

Now, let us assume the scenario where the receiver is receiving one data stream and can only feed back quantized information to the base station. Since $L_1 = 1$ in this case, $\mathbf{V}_1$ and $\mathbf{U}_1$ take the form of a vector. Let us represent those with $\mathbf{v}_1$ and $\mathbf{u}_1$ respectively. The receiver projects the MIMO channel $\mathbf{H}_1$ to an effective vector downlink channel $\mathbf{f}_1$ using $\mathbf{v}_1$ i.e., $\mathbf{f}_1 = \mathbf{H}_1\mathbf{v}_1$. Let $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N]$ be the quantization codebook and $\hat{\mathbf{f}}_1$ be the quantized effective vector downlink channel. Using, $\mathbf{u}_1 = \hat{\mathbf{f}}_1$, the signal power at the receiver becomes, $|\hat{\mathbf{f}}_1^H \mathbf{H}_1 \mathbf{v}_1|^2$. Using MET, with a large number of codevectors, the optimal solution takes the form [29],

$$\mathbf{v}_1 = \mathbf{B}_1(:, 1) \tag{2.22}$$

$$\hat{\mathbf{f}}_1 = \arg\max_{\mathbf{c}_i \in \mathbf{C}} |\langle \mathbf{c}_i, \mathbf{A}_1(:, 1) \rangle|^2 \tag{2.23}$$

Therefore, given a large number of codevectors, the optimal receive combining vector is the dominant right singular vector. The optimal quantized channel vector is the code-

vector that has maximum inner product with the dominant left singular vector. This algorithm leads to maximizing signal power.

### Quantization Based Combining (QBC)

Jindal introduced the idea of QBC [29]. In this section, we will show the motivation and description of QBC.

**Motivation** In this section, we use the description of Boccardi et al. [58] to prove the significance of QBC.

Let us assume that a single base station, consisting of $K$ transmit antennas, is communicating with $K$ users that are equipped with multiple receiver antennas. Let $P_{max}$ be the power budget. Each user receives one data stream. Therefore, $\mathbf{U}$ and $\mathbf{V}$ take the form of vectors. Let $\mathbf{H}_1, \cdots, \mathbf{H}_K$ be the channels to the $K$ users. $\mathbf{u}_k$ and $\mathbf{v}_k$ denote the precoding and decoding vectors of the k$^{th}$ stream. Let us assume $\mathbf{f}_k = \mathbf{H}_k \mathbf{v}_k$.

The SINR at the receiving end of the $k^{\text{th}}$ data takes the following form,

$$SINR_k = \frac{p_k |\mathbf{f}_k \mathbf{u}_k|^2}{\sigma^2 + \sum_{j \in [1,K], j \neq k} p_j |\mathbf{f}_k \mathbf{u}_j|^2} \tag{2.24}$$

Here, $\sigma^2$ is the noise variance and $p_k$ is the allocated power to the $k^{\text{th}}$ user. Let us assume that the receivers use unit norm shape feedback based on chordal distance. Therefore, the receivers choose the quantized effective channel, $\hat{\mathbf{f}}_k$, in the following way:

$$\hat{\mathbf{f}}_k = \arg \min_{\mathbf{c}_k \in \mathbf{C}} \sin^2 \left( \angle \left( \overline{\mathbf{f}}_k, \mathbf{c}_k \right) \right) \tag{2.25}$$

Here, $\overline{\mathbf{f}}_k = \mathbf{f}_k / ||\mathbf{f}_k||$. Let us define the quantization angle, $\theta_k$, and quantization error, $\tilde{\mathbf{f}}_k$ as,

$$\cos \theta_k = \left| \overline{\mathbf{f}}_k^H \hat{\mathbf{f}}_k \right| \tag{2.26}$$

$$\widetilde{\mathbf{f}}_k = \overline{\mathbf{f}}_k - \left( \widehat{\mathbf{f}}_k^H \overline{\mathbf{f}}_k \right) \widehat{\mathbf{f}}_k \tag{2.27}$$

It can be easily verified that, $||\widetilde{\mathbf{f}}_k||^2 = \sin^2\theta_k$. Using (2.27) in (2.24),

$$SINR_k = \frac{p_k||\mathbf{f}_k||^2 \left|\left|\left(\left(\hat{\mathbf{f}}_k^H\overline{\mathbf{f}}_k\right)\mathbf{f}_k + \tilde{\mathbf{f}}_k\right)^H \mathbf{u}_k\right|\right|^2}{\sigma^2 + ||\mathbf{f}_k||^2\sum_{j\in[1,K],j\neq k} p_j \left|\left|\left(\left(\hat{\mathbf{f}}_k^H\overline{\mathbf{f}}_k\right)\hat{\mathbf{f}}_k + \tilde{\mathbf{f}}_k\right)^H \mathbf{u}_j\right|\right|^2} \tag{2.28}$$

Let us assume that the BS uses zero-forcing precoder [30]. To calculate (2.28), the users make the following simplistic assumption: $\hat{\mathbf{f}}_i^H\hat{\mathbf{f}}_j = 0\,\forall\, i \neq j$ i.e. quantized effective channels of two different users are assumed to be multually orthogonal. Note that, this assumption is suboptimal since streams of two different users are only statistically independent in reality. However, this suboptimal assumption is considered due to the following two reasons:

1. The receivers do not cooperate with each other and therefore each user does not have access to other's channel knowledge.

2. The users find the receive combining vector $\mathbf{v}_i$ and design $\mathbf{f}_i$ prior to the design of $\mathbf{u}_i$.

The significance of this suboptimal assumption will be shown later in this section. Using these assumptions, $\mathbf{u}_k = \hat{\mathbf{f}}_k$, $\mathbf{u}_j^H\hat{\mathbf{f}}_k = 0$. Also, for high bit quantization, $\tilde{\mathbf{f}}_k^H\mathbf{u}_k \approx 0$. Therefore, (2.28) takes the following form,

$$\begin{aligned} SINR_k &= \frac{p_k||\mathbf{f}_k||^2\left|\hat{\mathbf{f}}_k^H\overline{\mathbf{f}}_k\right|^2}{\sigma^2 + ||\mathbf{f}_k||^2\sum_{j\in[1,K],j\neq k} p_j\left|\left|\tilde{\mathbf{f}}_k^H\mathbf{u}_j\right|\right|^2} \\ SINR_k &= \frac{\frac{P_{max}}{L}||\mathbf{f}_k||^2\cos^2_{\theta_k}}{\sigma^2 + \frac{P_{max}}{L}||\mathbf{f}_k||^2||\tilde{\mathbf{f}}_k||^2\sum_{j\in[1,K],j\neq k}\left|\left|\tilde{\overline{\mathbf{f}}}_k^H\mathbf{u}_j\right|\right|^2} \end{aligned} \tag{2.29}$$

Since power allocation across the users takes place after receive combining, we assume equal power allocation in (2.29). We also assume, $\tilde{\overline{\mathbf{f}}}_k = \frac{\tilde{\mathbf{f}}_k}{||\tilde{\mathbf{f}}_k||}$. Now, $E\left[\left|\left|\tilde{\overline{\mathbf{f}}}_k^H\mathbf{u}_j\right|\right|^2\right] = \frac{1}{M-1}$ [5]. Since (2.29) is convex with respect to $\sum_{j\in[1,K],j\neq k}\left|\left|\tilde{\overline{\mathbf{f}}}_k^H\mathbf{u}_j\right|\right|^2$, using Jensen's inequality [7],

$$E[SINR_k] \geq \frac{\frac{P_{max}}{L}||\mathbf{f}_k||^2\cos^2_{\theta_k}}{\sigma^2 + \frac{P_{max}}{L}||\mathbf{f}_k||^2\sin^2_{\theta_k}} \tag{2.30}$$

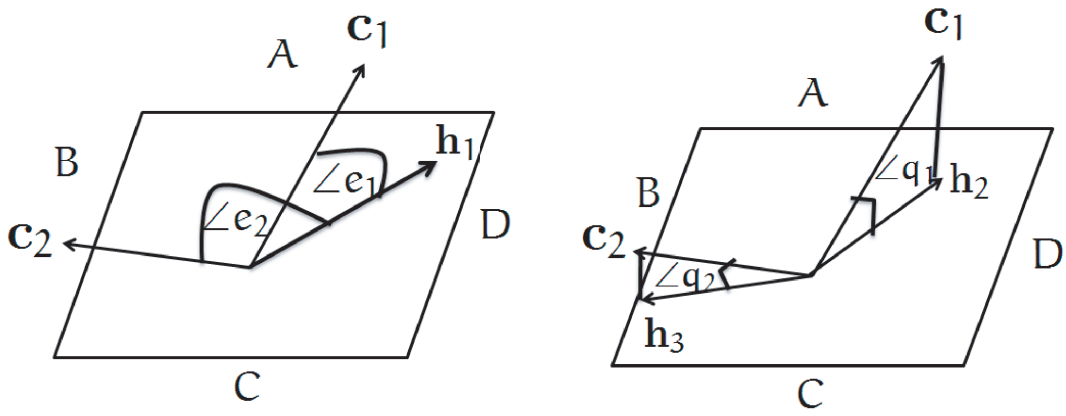



Figure 2.3: Comparison between Eigen-Based Combining and Quantization-Based Combining

At high signal-to-noise ratio, $\frac{P_{max}}{L}$ dominates over $\sigma^2$. So, $E[SINR_k] \geq \cot^2_{\theta_k}$. Therefore, minimization of $\theta_k$ leads to maximizing $SINR_k$. Thus, reduction of quantization error should be prioritized over increasing signal power in the high SNR regime of a multiuser system. This leads to the introduction of QBC in multiuser systems.

**Description of QBC** Similar to the sections described above, we assume a codebook $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_N]$ and the effective channel, $\mathbf{f}_k = \mathbf{H}_k \mathbf{v}_k$. [29] proposes to choose $\hat{\mathbf{f}}_k$ in the following way,

$$\hat{\mathbf{f}}_k = arg \min_{\mathbf{c}=\mathbf{c}_1,\cdots,\mathbf{c}_N} |\angle(\mathbf{c}, \mathbf{H}_k)| \tag{2.31}$$

Thus, the chosen codevector has the least quantization error with the span of the MIMO channel.

Fig. 2.3 clarifies the difference between quantization based combining and eigenbased combining. Here, $K = 1$, $M = 3$, $N = 2$, $L = 1$, $B = 1$. Having perfect channel estimation, the receiver has to execute two functions:

1. Combine the $3 \times 2$ MIMO channel into an effective $3 \times 1$ MISO channel.
2. Find the codevector that best represents the effective MISO channel.

In Fig. 2.3, $ABCD$ represents the subspace spanned by the columns of the MIMO

channel; *ox*,*oy* and *oz* denote the three axes. In this model, we assume 1 bit feedback overhead for simplicity. The codebook $\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2]$ where $\mathbf{c}_1$ and $\mathbf{c}_2$ are the two codevectors. The left and right side figures denote the case of eigenbased combining and quantization based combining respectively. In the left figure, $\mathbf{h}_1$ denotes the dominant eigenvector of the MIMO channel. On the other hand, $\mathbf{h}_2$ and $\mathbf{h}_3$ represent the projections of $\mathbf{c}_1$ and $\mathbf{c}_2$ in the subspace. Here, $\angle e_1 = \angle(\mathbf{c}_1, \mathbf{h}_1)$, $\angle e_2 = \angle(\mathbf{c}_2, \mathbf{h}_1)$, $\angle q_1 = \angle(\mathbf{c}_1, \mathbf{h}_2)$, $\angle q_2 = \angle(\mathbf{c}_2, \mathbf{h}_3)$.

In Eigenbased combining, the dominant eigenvector $\mathbf{h}_1$ will be the effective MISO channel. Since $\angle e_1 < \angle e_2$, $\mathbf{c}_1$ will be used as the quantized channel. On the other hand, in the right hand figure, $\angle q_1 > \angle q_2$ i.e., $\mathbf{c}_2$ is closer to the subspace. Therefore, in quantization based combining, $\mathbf{c}_2$ and $\mathbf{h}_3$ will be used as the quantized channel and effective MISO channel respectively.

### Maximum Expected Signal Combining

MET is the optimum approach in single user wireless communications. Using (2.24), it can be easily shown that it is optimal even in a broadcast channel at low SNR. On the other hand, QBC is optimal in a broadcast channel at high SNR. Trivellato et. al. [58] has introduced the maximum expected signal combining (MESC) algorithm that retains the benefits of MET and QBC at low and high SNR respectively. Since our proposed algorithm converges to MESC for one data stream, we skip the description of [58] here.

All these schemes discussed so far assume that each user receives a single data stream. However, our intent here is the general case wherein a user, with multiple receive antennas may receive multiple data streams [32, 55]. Multiple data streams per user complicates the feedback process, requiring independent information for each stream. In this part of our work, we extend the MESC algorithm to the multiple data stream scenario. Our contributions in this part of the work are given below:

1. We provide an end to end SMSE transceiver design that eliminates the dimension-

ality constraint and tie the feedback overhead to the number of data streams, which is always less than or equal to both the number of transmit and receive antennas.

2. We extend and make the MESC receiver flexible, by allowing for multiple data streams per user scenario.

3. We show why SMSE and BER increase, instead of converging to an error floor, if quantization error is not considered.

This concludes our literature review on quantization and receive combining schemes.

## 2.5 Bit Allocation across gain and shape feedback

Power allocation and interference cancellation across the users are both significant in a broadcast channel. The base station needs to be aware of both channel quality indicator (CQI) and channel shape indicator (CDI) to perform the two mentioned operations. In SQ, the real and imaginary entries of the channel coefficients are quantized separately and fed back to the base station. This helps the BS to estimate the CQI and CDI of the channel. However, in a VQ scenario, the quality of CQI and CDI depends on the number of bits allocated on the gain and shape of the channel vector. Therefore, optimal bit allocation across gain and shape feedback at the receiver end can become important in a limited feedback scenario.

To the best of our knowledge, there have not been many works on optimal bit allocation across gain and shape of the vector. The earliest work on this field can be traced to Hamkins & Zeger [20]. The authors assume a product codebook to quantize a vector in $\mathbb{R}^M$. Let, $\mathbf{x} = g\mathbf{s}$. Here, let $\mathbf{x}$ be the original channel. Here, $g$ and $\mathbf{s}$ represent the gain and shape of the channel respectively. Based on Euclidean distance as the distance metric, the authors use different codebooks to represent gain and shape separately. We skip the modeling of Hamkins' work due to its similarity with our proposed model that will be presented in chapter 4.

Khosnevis and Yu [33, 34] also worked on optimal bit allocation using product codebook. They use chordal distance as the distance metric. The authors consider a multiuser MISO system with real channels and try to minimize the total power allocated across the users subjects to outage constraints [34]. The authors provide optimal power and feedback bit allocation across the users and optimal bit allocation across the gain and shape of individual users.

In spite of the difference in objectives, the works in [20, 34] converge to almost similar results in terms of bit rate. Let $B$, $B_s$ and $B_g$ represent the overall bit, shape bit and gain bit respectively. Let, $B_s$ and $B_g$ represent the shape bit rate and gain bit rate respectively. Both [20] and [34] prove,

$$B_s = \frac{M-1}{M}B + k_1 \tag{2.32}$$

$$B_g = \frac{1}{M}B + k_2 \tag{2.33}$$

Here, $k_1$ and $k_2$ are constants with respect to bit allocation i.e., they depend on $M$, not $B$. This indicates the following result. If $M$ bits are available to quantize a MISO vector channel, approximately 1 and $(M-1)$ bits should be used to quantize the gain and shape respectively. The intuitive exlanation for this is as follows: the gain of a $\mathbb{R}^M$ vector follows a one dimensional distribution, whereas the shape lies uniformly in the $(M-1)$ dimensional surface of a unit norm hypersphere. As we will show in Chapter 4, our derivations also lead to a similar result in optimal bit allocation.

Since we consider a multiuser MIMO channel, quantization bits should be allocated optimally across gain and shape of the effective MISO channel (i.e., the effective vector downlink channel of the data stream that the receiver obtains after performing receive combining on the MIMO channel). The shape of the effective MISO channel is uniformly distributed in $\mathbb{C}^M$. However, the norm of the effective MISO channel depends on the receive combining algorithm. In EBC, the norm of the effective MISO channel takes the form of the dominant singular value of the MIMO matrix. The distribution of the

eigenvalues of the Wishart matrix is given in [54]. On the other hand, the distribution of the norm of the effective MISO channel in QBC follows a chi-squared distribution with $(M - N_k + 1)$ [30].

Since MESC converges to EBC at low SNR and QBC at high SNR, to the best of our knowledge, there has not been any work that provides a general expression of the distribution of the norm of the effective MISO channel in MESC combining. Therefore, albeit being non-optimal at high SNR broadcast channels, we use EBC in our work on optimal bit allocation.

The difference between our work and the works in [20, 34] can be summarized as follows:

1. We provide optimal bit allocation across the eigenvalue and eigenvector of a MIMO channel.

2. Due to the use of SMSE precoder, we use Euclidean distance as the distance metric while assuming complex channels.

## 2.6 Channel Quantization in time varying channels

So far our discussion has focused on block fading channels. In the last part of our work, we focus on channel quantization in time varying channels. Supporting mobility is an integral part of next generation broadband wireless networks [2]. The CSI overhead in a MIMO channel can be significantly reduced using differential feedback by exploiting temporal correlation of the channel [24, 35, 36]. Most of these works assume the channel as a first-order Gauss-Markov process. Assuming a single input single output channel (SISO) , the first order Gauss-Markov process can be illustrated as follows,

$$h(n) = a \times h(n-1) + \sqrt{1-a^2}\zeta(n) \tag{2.34}$$

Here, $h(n)$ is the the SISO channel at the $n^{\text{th}}$ instant. $a$ is the temporal correlation between two successive channel samples and $\zeta(n)$ is a white innovation process independent

of $h(n)$, i.e.,

$$E\left[h(n)h(n-1)\right] = a \tag{2.35}$$

$$E\left[h(n)\zeta(n)\right] = 0 \tag{2.36}$$

Both $h(n)$ and $\zeta(n)$ follow a Gaussian distribution with same variance. The authors assume that both transmitter and receiver are aware of $a$. Let, $\hat{h}(n)$ and $\hat{h}(n-1)$ be the quantized versions of $h(n)$ and $h(n-1)$ respectively. The authors in [24, 35, 36] propose the following feedback model,

$$\hat{h}(n) = a\hat{h}(n-1) + \hat{d}(n) \tag{2.37}$$

Here, $\hat{d}(n)$ is the quantized version of the difference signal, $d(n) = h(n) - h(n-1)$. The temporal correlation, $a \neq 0$, is assumed to lie between 0 and 1 and can be calculated through the Doppler fading process.

The model shown above has the following limitations:

1. The autocorrelation curve obtained from Markov chain model differs from the well-accepted Jakes' model significantly when normalized autocorrelation between two successive sample drops below 0.5 (approximately 16/17 Km/hr in present wireless communication standards) [17, 24].

2. The works in [24,35,36] assume that the transmitter and receiver agree on the value of the parameters in the Markov chain. This assumption does not hold in non-stationary channels.

There have been some works in the literature that avoid the two limitations mentioned above. The authors in these works mostly focus on adaptive delta modulation based feedback [47, 56]. These works are based on Jayant's work on ADM in speech coding [28] and quantize the difference between the previous and current samples with a one-bit quantizer. This form of delta modulation uses an individual delta modulator to track the real and imaginary entries of each of the real and imaginary parts of the channel coefficient. The step size of the delta modulator is adapted according to $h(n)$ and $h(n-1)$

respectively. If the previous two encoded bits are same, the step size is increased by a factor of $\alpha$ and vice versa. In spite of the apparent advantage of the algorithm, the authors only provide suitable step size parameters for pedestrian velocities (e.g., 4 km/hr).

The lack of flexibility of the proposed differential feedback methods motivates us to investigate adaptive differential feedback in time-varying multiuser channels. We make the following two contributions in this work:

1. Based on the model of adaptive differential modulation model proposed by Stroh [53], we develop 2-bit recursive least square (RLS) and linear least square (LLS) based adaptive differential feedback methods in a time-varying environment. We show that 2-bit adaptive differential feedback can outperform 3-bit and 2-bit fixed feedback up to 18 km/h and up to 32 km/h respectively.

2. We design a RLS adaptive tracking of the eigenvectors of each user's channel matrix and show that, if the number of data streams is less than the total number of receive antennas, this method reduces feedback overhead.

Both these methods can lead to reducing the required feedback by several kbits/sec in modern wireless communication standards.

# Chapter 3

# Precoder Design with Limited Feedback in a Block Fading Channel

In this chapter, we propose a solution for the SMSE minimization of a multiuser MIMO system with limited feedback. The contribution of this chapter is two-fold. First, we provide an end-to-end SMSE transceiver design that incorporates receiver combining, feedback policy and transmit precoder design with channel uncertainty. Second, we remove dimensionality constraints on the MIMO system, for the scenario with multiple data streams per user, using a combination of maximum expected signal combining and minimum MSE receiver. This makes each user's feedback independent of the others and the resulting feedback overhead scales linearly with the number of data streams instead of the number of receiving antennas.

## 3.1 Problem Formulation

The system model (both downlink and uplink) used here was presented in the previous chapter. Therefore, Fig. 2.1 and Fig. 2.2 still represent the downlink and uplink system model. In this chapter, we at first reiterate the original objective. Thereafter, we show how the multiuser MIMO problem can be transformed into an equivalent MU MISO

problem.

From section 2.1.1, the estimated data of user $k$ is found as follows,

$$\hat{\mathbf{x}}_k = \mathbf{V}_k^H \mathbf{y}_k^{DL} \tag{3.1}$$

$$= \mathbf{V}_k^H \mathbf{H}_k^H \mathbf{U}\sqrt{\mathbf{P}}\mathbf{x} + \mathbf{V}_k^H \mathbf{n}_k \tag{3.2}$$

Assuming $\mathbf{V}$ to be the block diagonal global decoder matrix,

$$\mathbf{V} = diag\left(\mathbf{V}_1, ..., \mathbf{V}_K\right) \tag{3.3}$$

Combining all users,

$$\mathbf{y} = \mathbf{H}^H \mathbf{U}\sqrt{\mathbf{P}}\mathbf{x} + \mathbf{n} \tag{3.4}$$

$$\hat{\mathbf{x}} = \mathbf{V}^H \mathbf{H}^H \mathbf{U}\sqrt{\mathbf{P}}\mathbf{x} + \mathbf{V}^H \mathbf{n} \tag{3.5}$$

where, $\mathbf{n} = \left[\mathbf{n}_1^T, \mathbf{n}_2^T, \ldots, \mathbf{n}_K^T\right]^T$.

$\mathbf{E}_k^{DL}$ denotes the $L_k \times L_k$ error covariance matrix of user $k$ in the downlink, where

$$\mathbf{E}_k^{DL} = E\left[\left(\hat{\mathbf{x}}_k - \mathbf{x}_k\right)\left(\hat{\mathbf{x}}_k - \mathbf{x}_k\right)^H\right] \tag{3.6}$$

The diagonal entries of $\mathbf{E}_k^{DL}$ are the MSEs of the $L_k$ substreams of user $k$. Therefore, $SMSE_k^{DL} = tr\left[\mathbf{E}_k^{DL}\right]$, where $tr[\cdot]$ denotes the trace operator. The SMSE minimization problem can be formulated as,

$$min_{\mathbf{p},\mathbf{U},\mathbf{V}} \sum_{k=1}^{K} tr\left[E_k^{DL}\right] \tag{3.7}$$

$$subject\,to: ||\mathbf{p}||_1 \leq P_{max}$$

## 3.2 Transformation to a Equivalent MU MISO System

The SMSE minimization problem in (3.7) depends on the power allocation matrix $\mathbf{P}$, beamformer $\mathbf{U}$ and downlink decoder $\mathbf{V}$. Since users cannot cooperate with each other,

the joint design of $\mathbf{P}$, $\mathbf{U}$ and $\mathbf{V}$ has to take place at the base station. This requires the presence of complete channel knowledge at the base station. The required feedback overhead to send back the full channel information i.e., $\mathbf{H}$, sclaes as $M \times N$. Here, $M$ and $N$ denote the total number of transmit and receive antennas of the system.

We propose a *sub-optimal alternative to this joint design* in our work. Instead of sending back the full channel knowledge, each user can design the receive decoding vector in its own end and then feed back the matrix $\mathbf{HV}$ to the base station. In this case, the feedback overhead will scale with $M \times L$. Here, $L$ denotes the total number of data streams that the users receive. Since $L \leq N$, *this method minimizes the feedback overhead, at the cost of a suboptimal* $\mathbf{V}$. To the best of our knowledge, all the existing works on receive combining use maximizing SINR as the design criterion. We also follow the same approach in our work.

Now, in (3.5), let us assume $\mathbf{F} = \mathbf{HV}$. Here $F$ is a $M \times L$ that is formulated as follows,

$$\begin{aligned} \mathbf{F} &= [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_L] && (3.8) \\ &= \left[\mathbf{f}_{1_1}, \mathbf{f}_{1_2}, \cdots, \mathbf{f}_{1_{L_1}}, \mathbf{f}_{2_1}, \mathbf{f}_{2_2}, \cdots, \mathbf{f}_{2_{L_2}}, \cdots, \mathbf{f}_{K_1}, \cdots, \mathbf{f}_{K_{L_K}}\right] && (3.9) \\ &= [\mathbf{F}_1, \mathbf{F}_2, ..., \mathbf{F}_K] && (3.10) \end{aligned}$$

$\mathbf{f}_{k_j} = \mathbf{H}_k \mathbf{v}_{k_j}$ is a length-M vector that denotes the effective vector downlink channel of the j$^{th}$ stream of the k$^{th}$ user. Here $\mathbf{H}_k$ is the channel of the $k^{\text{th}}$ user and $\mathbf{v}_{k_j}$ is the decoding vector used for its $j^{\text{th}}$ stream. Essentially, the $k^{\text{th}}$ user combines its $H_k$ channel to $\mathbf{f}_{k_j}$ using the decoding vector $\mathbf{v}_{k_j}$ resulting in a set of projected MISO channels for each of its data streams. Thus, the columns of $\mathbf{F}$ have become the effective downlink MISO channels of the data streams of the whole system.

Using this transformation, the overall system equation takes the following form:

$$\hat{\mathbf{x}} = \mathbf{F}^H \mathbf{U} \sqrt{\mathbf{P}} \mathbf{x} + \mathbf{V}^H \mathbf{n} \tag{3.11}$$

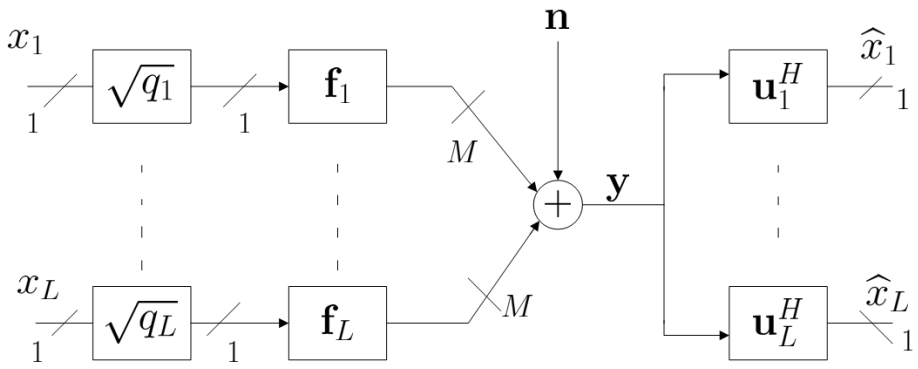



Figure 3.1: Block diagram of Multiuser MIMO uplink with channel and decoder combined as a single block

Now, the equations in the virtual uplink takes the following form,

$$\mathbf{y}^{UL} = \left( \sum_{j=1}^{L} \mathbf{f}_j \sqrt{q_j} x_j + \mathbf{n} \right) \tag{3.12}$$

$$\hat{x}_i^{UL} = \mathbf{u}_i^H \left( \sum_{j=1}^{L} \mathbf{f}_j \sqrt{q_j} x_j + \mathbf{n} \right) \tag{3.13}$$

Fig. 3.1 shows the proposed system model in the virtual uplink. As (3.13) and Fig. 3.1 show, the system has become an effective MU MISO system. The uplink-downlink duality in a multiuser MISO system under imperfect channel conditions has already been proved in [14].

Therefore, the new design objective becomes a two-step design problem.

*Design at the receiver end:*

$$max_{\mathbf{V}_k} \mathbf{E}\left[SINR_k\right] \tag{3.14}$$

*Design at the base station:*

$$min_{\mathbf{p},\mathbf{U}} \sum_{k=1}^{K} tr\left[E_k^{UL}\right] \tag{3.15}$$

$$subject\,to: ||\mathbf{p}||_1 \leq P_{max}$$

Here, $\mathbf{E}\left[SINR_k\right]$ denotes the expected Signal-to-interference-plus-noise-ratio (SINR) at the receiver. Since users cannot cooperate with each other, each user has to find the

decoding vector on its own prior to the design of the precoding vector and power allocator. Therefore, the users can only find the expected SINR at the transmitter.

## 3.3 Overall Algorithm

We, at first, summarize the overall algorithm. Important steps of the algorithm will be described in details in the following section.

1. Send common pilots to the users in the system so that each user can estimate its own channel.

2. In the MU MIMO case, each user converts its estimated MIMO channel to effective MISO channels using the MESC algorithm, proposed in Section 3.4.2. After MSIP based quantization, each user sends the codebook indexes of the effective channels to the BS. In a MU MISO system, each user quantizes its own channel and the BS assumes $\mathbf{V} = \mathbf{I}$.

3. Virtual uplink power allocation is solved using a convex optimization problem formulation, shown in Section 3.5.

4. Uplink beamformer takes the form of a weiner filter, shown in Section 3.5.

5. Downlink power allocation $\mathbf{P} = \mathbf{Q}$.

6. Send dedicated pilot symbols for each of the data streams. Thereafter, implement the MMSE downlink decoders, shown in Section 3.6.

## 3.4 Receiver End Design after Channel Estimation

We propose a two step receiver design. For the purposes *of quantization only*, each user uses a MESC receiver and chooses the quantized codevectors that would maximize

the SINR of their data streams. However, the users implement MMSE receivers while receiving the actual data. This allows for the *channel feedback to be independent of the other users' actions.*

We will use the symbol policy of (3.9) in this section. We assume that the receivers have perfect CSI using training, i.e., the k$^{th}$ user knows its channel $\mathbf{H}_k$ completely. Now, for each of its receiving data stream $k_j$ where $j \in [1, \cdots, L_k]$, each user has to execute two operations:

1. Generate a codebook $\mathbf{C}$ consisting of $2^B$ unit norm vectors $\mathbf{c}_1, ..., \mathbf{c}_{2^B}$ off-line.

2. Design the decoding filter $\mathbf{v}_{k_j}$ for its j$^{th}$ stream and find the effective vector downlink channel $\mathbf{f}_{k_j}$ using $\mathbf{f}_{k_j} = \mathbf{H}_k \mathbf{v}_{k_j}$.

3. Quantize the original effective vector downlink channel $\mathbf{f}_{k_j}$ to a quantized vector downlink channel $\hat{\mathbf{f}}_{k_j}$.

### 3.4.1 Codebook Generation and Quantization

Each user feeds back $B$ bits per data stream to the BS. The k$^{th}$ user quantizes $\mathbf{f}_{k_j}$ using the chordal distance [10]:

$$\hat{\mathbf{f}}_{k_j} = \arg \min_{\mathbf{c} \in \left[\mathbf{c}_1, .., \mathbf{c}_{2^B}\right]} \sin^2 \left( \angle \left( \mathbf{f}_{k_j}, \mathbf{c} \right) \right) \tag{3.16}$$

The use of chordal distance over the Euclidean distance leads to a higher inner product between the original and quantized channels [48]. Here, we only quantize the direction of the effective channel and this direction can lie anywhere on the $M$-dimensional complex unit-norm sphere. Therefore, we generate the quantization codebook as a VQ problem using the MSIP optimality criterion; the details of MSIP VQ codebook generation can be found in Appendix A.

### 3.4.2 Receive Combiner Design

Using our quantization policy in (3.16), we define the quantization angle $\theta_{k_j} \in [0, \frac{\pi}{2}]$ as,

$$\cos\theta_{k_j} = \left|\overline{\mathbf{f}}_{k_j}^H \hat{\mathbf{f}}_{k_j}\right| \tag{3.17}$$

Here, $\overline{\mathbf{f}}_{k_j} = \frac{\mathbf{f}_{k_j}}{||\mathbf{f}_{k_j}||_2}$.

Since, the receivers know the quantization angle exactly, we can use this information to improve the expected SINR. As in [58], define the quantization error as,

$$\widetilde{\mathbf{f}}_{k_j} = \overline{\mathbf{f}}_{k_j} - \left(\widehat{\mathbf{f}}_{k_j}^H \overline{\mathbf{f}}_{k_j}\right) \widehat{\mathbf{f}}_{k_j} \tag{3.18}$$

Here, $||\widetilde{\mathbf{f}}_{k_j}||^2 = \sin^2\theta_{k_j}$. Now, in the downlink, the SINR of the $j^{\text{th}}$ stream of the $k^{th}$ user is,

$$\text{SINR}_{k_j}^{DL} = \frac{\frac{P}{L}\left|\mathbf{f}_{k_j}^H \mathbf{u}_{k_j}\right|^2}{\sigma^2 + \sum_{n\neq j}\frac{P}{L}\left|\mathbf{f}_{k_j}^H \mathbf{u}_{k_n}\right|^2 + \sum_{m\in[1,K],m\neq k,l\in[1,L_m]}\frac{P}{L}\left|\mathbf{f}_{k_j}^H \mathbf{u}_{m_l}\right|^2} \tag{3.19}$$

In (3.19) equal power allocation was assumed to simpify the receiver combining analysis. Here, $\sum_{n\neq j}\frac{P}{L}\left|\mathbf{f}_{k_j}^H \mathbf{u}_{k_n}\right|^2$ and $\sum_{m\in[1,K],m\neq k,l\in[1,L_m]}\frac{P}{L}\left|\mathbf{f}_{k_j}^H \mathbf{u}_{m_l}\right|^2$ denote intra-user and inter-user interference respectively.

We design transmit filters in section 3.5. Now using the solution of $\mathbf{u}_{k_j}$ ((3.47) and (3.48)) and the matrix inversion lemma,

$$\begin{aligned}\mathbf{u}_{k_j} &= \left(\left(\sigma^2 + \frac{\sigma_E^2}{M}P_{max}\right)\mathbf{I} + \widehat{\mathbf{F}}\mathbf{Q}\widehat{\mathbf{F}}^H\right)^{-1}\hat{\mathbf{f}}_{k_j}\sqrt{q_{k_j}} \\ &= \frac{1}{\sigma^2 + \frac{\sigma_E^2}{M}P_{max}}\left[\mathbf{I} - \frac{1}{\sigma^2 + \frac{\sigma_E^2}{M}P_{max}}\hat{\mathbf{F}}\left(\mathbf{Q}^{-1} + \frac{1}{\sigma^2 + \frac{\sigma_E^2}{M}P_{max}}\hat{\mathbf{F}}^H\hat{\mathbf{F}}\right)^{-1}\hat{\mathbf{F}}^H\right]\hat{\mathbf{f}}_{k_j}\sqrt{q}_{k_j}\end{aligned} \tag{3.20}$$

Here, $\mathbf{u}_{k_j}$ is normalized such that, $||\mathbf{u}_{k_j}|| = 1$.

#### Simplistic approximations and brief preview of the receive combiner design

Now, (3.19) shows that the SINR of the $j^{\text{th}}$ stream of the $k^{\text{th}}$ user depends on the precoding vectors of all other data streams. However, as (3.20) suggests, precoding vector

of each stream depends on the effective vector downlink channel of all other streams. The receiver combiner design, i.e., the design of $\mathbf{v}_{k_j}$ and $\mathbf{f}_{k_j}$ takes place prior to the transmit filter design. Since we do not assume co-operation among the users, it is impossible for the receivers to find the exact value of (3.19). Therefore, we make some simplistic approximations in our design. Here, we summarize the overall receive combiner design and show the simplistic approximations,

1. We approximate that the effective vector downlink channels of the data streams of two different users are mutually orthogonal. Therefore, $\hat{\mathbf{f}}_{k_i}^H \hat{\mathbf{f}}_{l_j}^H = 0$ where $k \in [1, K], l \in [1, K], k \neq l$, $i \in [1, \cdots, Ł_k]$, $j \in [1, \cdots, L_l]$.

2. While designing the decoding filter of its $1^{\text{st}}$ data stream, $k^{\text{th}}$ user assumes the quantized effective vector downlink channel of this stream to be orthogonal to that of its all other data streams. Therefore, while designing $\mathbf{v}_{k_1}$, the $k^{\text{th}}$ user approximates, $\hat{\mathbf{f}}_{k_1}^H \hat{\mathbf{f}}_{k_j} = 0, j \in [2, \cdots, L_k]$. Based on this approximation, the $k^{\text{th}}$ user finds the $\mathbf{f}_{k_1}$ and $\mathbf{v}_{k_1}$ that maximizes $SINR_{k_1}$.

3. While designing the decoding filter of its $2^{\text{nd}}$ data stream, $k^{\text{th}}$ user approximates the quantized effective vector downlink channel of this stream to be orthogonal to that of its other data streams whose decoding filters and quantized channels have not been designed. Therefore, while designing $\mathbf{v}_{k_2}$, the $k^{\text{th}}$ user approximates, $\hat{\mathbf{f}}_{k_2}^H \hat{\mathbf{f}}_{k_j} = 0, j \in [3, \cdots, L_k]$. Based on this approximation, the $k^{\text{th}}$ user finds the $\mathbf{f}_{k_2}$ and $\mathbf{v}_{k_2}$ that maximizes $SINR_{k_2}$.

4. The $k^{\text{th}}$ user continues the same policy up to its $L_k^{\text{th}}$ stream.

Note that, these simplistic approximations are considered only at the receive combiner design due to lack of knowledge of the other user's channels. Since the BS has access to the effective vector downlink channel of all the data streams, the BS does not consider these simplistic assumptions and find the transmit filter to minimize the overall SMSE.

**Detailed descriptions of the approximations and the receive combiner design**

Using the approximation of orthogonal channel of two different user's streams, it can be easily verified from (3.20) that $\hat{\mathbf{f}}_{k_i}^H \hat{\mathbf{u}}_{l_j}^H = 0$ where $k \in [1, K], l \in [1, K], k \neq l, i \in [1, \cdots, Ł_k]$, $j \in [1, \cdots, L_l]$.

Since each user knows the inner product of different code vectors in its codebook, the assumption of orthogonality is not valid for two different streams of the same user. Therefore, in our proposed algorithm, each user uses its known codevectors, i.e., the effective channels of its data streams, as a set of column vectors $\widehat{\mathbf{f}}$ in the $\widehat{\mathbf{F}}$ matrix and assumes that the vector downlink channels for all other users' stream are mutually orthogonal to its own channels. We also assume that noise variance, signal power, quantization error variance in the BS and total number of data streams sent by the BS are known to each of the users. Therefore due to the construction of (3.20), each user can approximate the expected value of $\mathbf{f}_{k_j}^H \mathbf{u}_{k_j}$ and $||\mathbf{u}_{k_j}||$ even without co-operating with other users. Now, from (3.19),

$$SINR_{k_j}^{DL} = \frac{\frac{P}{L}|\mathbf{v}_{k_j}^H \mathbf{H}_k^H \mathbf{u}_{k_j}|^2}{\sigma^2 + \sum_{n \neq j} \frac{P}{L} \left|\mathbf{f}_{k_j}^H \mathbf{u}_{k_n}\right|^2 + \sum_{m \in [1,K], m \neq k, l \in [1, L_m]} \frac{P}{L} \left|\mathbf{f}_{k_j}^H \mathbf{u}_{m_l}\right|^2} \tag{3.21}$$

Now,

$$\widetilde{\mathbf{f}}_{k_j} = \overline{\mathbf{f}}_{k_j} - \left(\widehat{\mathbf{f}}_{k_j}^H \overline{\mathbf{f}}_{k_j}\right) \widehat{\mathbf{f}}_{k_j} \tag{3.22}$$

$$\sum_{m \in [1,K], m \neq k, l \in [1, L_m]} \frac{P}{L} \left|\mathbf{f}_{k_j}^H \mathbf{u}_{m_l}\right|^2 = ||\mathbf{f}_{k_j}||^2 \sum_{m \in [1,K], m \neq k, l \in [1, L_m]} \left|\left|\left(\widehat{\mathbf{f}}_{k_j}^H \overline{\mathbf{f}}_{k_j}\right) \widehat{\mathbf{f}}_{k_j}^H \mathbf{u}_{m_l} + \widetilde{\mathbf{f}}_{k_j}^H \mathbf{u}_{m_l}\right|\right|^2 \tag{3.23}$$

$$= ||\mathbf{f}_{k_j}||^2 \sum_{m \in [1,K], m \neq k, l \in [1, L_m]} \left|\left|\widetilde{\mathbf{f}}_{k_j}^H \mathbf{u}_{m_l}\right|\right|^2 \tag{3.24}$$

$$= ||\mathbf{f}_{k_j}||^2 ||\widetilde{\mathbf{f}}_{k_j}||^2 \sum_{m \in [1,K], m \neq k, l \in [1, L_m]} \left|\left|\overline{\widetilde{\mathbf{f}}}_{k_j}^H \mathbf{u}_{m_l}\right|\right|^2 \tag{3.25}$$

$$= ||\mathbf{f}_{k_j}||^2 \sin^2 \theta_{k_j} \frac{L - L_k}{M - 1} \tag{3.26}$$

$$= \frac{L - L_k}{M - 1} \left(||\mathbf{f}_{k_j}||^2 - \left(\mathbf{f}_{k_j}^H \widehat{\mathbf{f}}_{k_j}\right) \left(\widehat{\mathbf{f}}_{k_j}^H \mathbf{f}_{k_j}\right)\right) \tag{3.27}$$

$$= \frac{L - L_k}{M-1} \mathbf{v}_{k_j}^H \left( \mathbf{H}_k^H \left( \mathbf{I} - \hat{\mathbf{f}}_{k_j} \hat{\mathbf{f}}_{k_j}^H \right) \mathbf{H}_k \right) \mathbf{v}_{k_j} \tag{3.28}$$

Here, (3.23) is obtained by using $\mathbf{f}_{k_j} = ||\mathbf{f}_{k_j}|| \overline{\mathbf{f}}_{k_j}$ and (3.22). (3.24) follows since we assume $\widehat{\mathbf{f}}_{k_j}^H \mathbf{u}_{m_l} = 0$ , $k \neq m$ for mutually orthogonal reported channels from different users. (3.25) was obtained by setting $\overline{\widetilde{\mathbf{f}}}_{k_j} = \frac{\widetilde{\mathbf{f}}_{k_j}}{||\widetilde{\mathbf{f}}_{k_j}||}$. (3.26) was derived using the analysis of [58]. In the presence of a large number of codevectors, $\theta_{k_j}$ is very small which leads to $\widetilde{\mathbf{f}}_{k_j}^H \widehat{\mathbf{f}}_{k_j} \approx 0$. Therefore, the unit vectors $\overline{\widetilde{\mathbf{f}}}_{k_j}$ and $\mathbf{u}_{m_l}$ are both identically distributed in the $(M-1)$ dimensional plane orthogonal to $\widehat{\mathbf{f}}_{k_j}$. This implies, $||\overline{\widetilde{\mathbf{f}}}_{k_j}^H \mathbf{u}_{m_l}||^2$ follows a beta distribution with parameter $(1, M-1)$ and has expected value $1/(M-1)$ [58]. The factor of $(L - L_k)$ arises since the $k^{\text{th}}$ user receives $L_k$ data streams and therefore $(L - L_k)$ data streams are mutually orthogonal to its $j^{\text{th}}$ data stream. (3.27) was obtained using the quantization angle definition of (3.17). In (3.28), we again use $\mathbf{f}_{k_j} = \mathbf{H}_k \mathbf{v}_{k_j}$.

Using the results of (3.28) in (3.21) and defining

$$\mathbf{B}_{k_j} = \frac{P}{L} \mathbf{H}_k^H \left( \sum_{n \in [1, L_k], n \neq j} \mathbf{u}_{k_n} \mathbf{u}_{k_n}^H + \frac{L - L_k}{M-1} \left( \mathbf{I} - \mathbf{f}_{k_j} \mathbf{f}_{k_j}^H \right) \right) \mathbf{H}_k \tag{3.29}$$

(3.21) takes the following form:

$$SINR_{k_j}^{DL} = \frac{\mathbf{v}_{k_j}^H \frac{P}{L} \mathbf{H}_k^H \mathbf{u}_{k_j} \mathbf{u}_{k_j}^H \mathbf{H}_k \mathbf{v}_{k_j}}{\sigma^2 + \mathbf{v}_{k_j}^H \mathbf{B}_{k_j} \mathbf{v}_{k_j}} \tag{3.30}$$

Due to the structure of $\mathbf{u}_{k_j}$ and $\mathbf{B}_{k_j}$, $SINR_{k_j}^{DL}$ in (3.30) is a function of $\mathbf{v}_{k_j}$ and $\hat{\mathbf{f}}_{k_j}$ $\forall j \in [1, L_k]$. Each of these $\hat{\mathbf{f}}_{k_j}$ vectors are in the codebook $\mathbf{C}$ which consists of codevectors $\mathbf{c}_1, \cdots, \mathbf{c}_{2^B}$. Therefore, the linear decoding vector $\mathbf{v}_{k_j}$ and $\mathbf{f}_{k_j} \forall j \in [1, L_k]$ is chosen as,

$$\left( \mathbf{f}_{k_j}, \mathbf{v}_{k_j} \right) \forall j \in [1, L_k] = \arg \max_{||\mathbf{v}_{k_j}||=1, \hat{\mathbf{f}}_{k_j} \in \mathbf{C}} SINR_{k_j}^{DL} \tag{3.31}$$

The detailed description of the algorithm is below:

1. First, assume that intra-user streams are orthogonal and find the vector downlink channel of the first stream. Therefore, maximizing (3.30) becomes an optimization problem of $\hat{\mathbf{f}}_{k_1}$ and $\mathbf{v}_{k_1}$. So,

$$\mathbf{B}_{k_1} = \left( \frac{P}{L} \mathbf{H}_k^H \left( \frac{L-1}{M-1} \left( \mathbf{I} - \hat{\mathbf{f}}_{k_1} \hat{\mathbf{f}}_{k_1}^H \right) \right) \mathbf{H}_k \right) \tag{3.32}$$

$$SINR_{k_1}^{DL} = \frac{\mathbf{v}_{k_1}^H \left(\frac{P}{L}\mathbf{H}_k^H \mathbf{u}_{k_1}\mathbf{u}_{k_1}^H \mathbf{H}_k\right)\mathbf{v}_{k_1}}{\sigma^2 + \mathbf{v}_{k_1}^H \mathbf{B}_{k_1}\mathbf{v}_{k_1}} \tag{3.33}$$

$$\left(\hat{\mathbf{f}}_{k_1}, \mathbf{v}_{k_1}\right) = \max_{\left(||\mathbf{v}_{k_1}||=1, \hat{\mathbf{f}}_{k_1}\in\mathbf{C}\right)} SINR_{k_1}^{DL} \tag{3.34}$$

2. Once the quantized channel of the 1st stream is chosen, the user assumes it to be a nonorthogonal channel for the second stream's vector downlink channel. However, vector downlink channels for the other streams of the same user are still considered to be orthogonal to both first and second stream's channel. Thus maximizing (3.30) again becomes an optimization problem with variables $\mathbf{v}_{k_2}$ and $\hat{\mathbf{f}}_{k_2}$ for the present data stream. So,

$$\mathbf{B}_{k_2} = \left(\frac{P}{L}\mathbf{H}_k^H \left(\mathbf{u}_{k_1}\mathbf{u}_{k_1}^H + \frac{L-2}{M-1}\left(\mathbf{I} - \hat{\mathbf{f}}_{k_2}\hat{\mathbf{f}}_{k_2}^H\right)\right)\mathbf{H}_k\right) \tag{3.35}$$

$$SINR_{k_2}^{DL} = \frac{\mathbf{v}_{k_2}^H \left(\frac{P}{L}\mathbf{H}_k^H \mathbf{u}_{k_2}\mathbf{u}_{k_2}^H \mathbf{H}_k\right)\mathbf{v}_{k_2}}{\sigma^2 + \mathbf{v}_{k_2}^H \mathbf{B}_{k_2}\mathbf{v}_{k_2}} \tag{3.36}$$

$$\left(\hat{\mathbf{f}}_{k_2}, \mathbf{v}_{k_2}\right) = \max_{\left(||\mathbf{v}_{k_2}||=1, \hat{\mathbf{f}}_{k_2}\in\mathbf{C}\right)} SINR_{k_2}^{DL} \tag{3.37}$$

3. For the 3rd data stream of the $k^{\text{th}}$ user,

$$\mathbf{B}_{k_3} = \left(\frac{P}{L}\mathbf{H}_k^H \left(\mathbf{u}_{k_1}\mathbf{u}_{k_1}^H + \mathbf{u}_{k_2}\mathbf{u}_{k_2}^H + \frac{L-3}{M-1}\left(\mathbf{I} - \hat{\mathbf{f}}_{k_3}\hat{\mathbf{f}}_{k_3}^H\right)\right)\mathbf{H}_k\right) \tag{3.38}$$

The other equations take the forms of (3.36) - (3.37). The same policy continues upto the last stream of the $k^{\text{th}}$ user.

With this algorithm, the SINR expression for a particular data stream remains a function of only its decoding vector and its quantized channel. This leads to a computational complexity of $L_k \times 2^B$ in finding the channels of $L_k$ data streams. Now (3.29) and (3.30) can be thought as a general form of all the data stream's SINR expressions. In (3.29) and (3.30), both $\mathbf{f}_{k_j}$ and $\mathbf{u}_{k_j}$ depend on the chosen codevector $\hat{\mathbf{f}}_{k_j}$. For any particular $\hat{\mathbf{f}}_{k_j}$, the linear decoding vector that maximizes (3.30) can be obtained by the MMSE detector, $\mathbf{v}_{k_j} = \left(\sigma^2\mathbf{I} + \mathbf{B}_{k_j}\right)^{-1}\sqrt{\frac{P}{L}}\mathbf{H}_k^H\mathbf{u}_{k_j}$ [58]. Then,

$$SINR_{k_j}^{DL} = \frac{P}{L}\mathbf{u}_{k_j}^H \mathbf{H}_k \left(\sigma^2\mathbf{I} + \mathbf{B}_{k_j}\right)^{-1}\mathbf{H}_k^H\mathbf{u}_{k_j} \tag{3.39}$$

The user finds the value of $SINR_{k_j}^{DL}$ for $\mathbf{c}_i \forall i \in [1, 2^B]$ using (3.39) and chooses the $\mathbf{c}_i$, as the quantized channel $\hat{\mathbf{f}}_{k_j}$, that maximizes $SINR_{k_j}^{DL}$.

It is worth emphasizing that, to our knowledge, this is the first receive combining scheme that considers signal power, inter-user and intra-user interference while accounting for multiple data streams per user.

## 3.5 Linear Precoder Design

The transmit precode design at the BS is performed from the data stream's perspective. Therefore, we use the symbol policy of (3.8) in this section. We consider the following channel model at the BS for precoder design,

$$\mathbf{f}_i = \widehat{\mathbf{f}}_i + \widetilde{\mathbf{f}}_i \quad \text{or} \quad \mathbf{F} = \widehat{\mathbf{F}} + \widetilde{\mathbf{F}} \tag{3.40}$$

Here, $\mathbf{f}_i$ denotes the effective vector downlink channel of the $i^{\text{th}}$ stream. $\mathbf{F}$ comprises $L$ unit-norm effective channel vectors with the original channel directions. $\hat{\mathbf{F}}$ denotes the $L$ quantized feedback unit norm vectors. $\tilde{\mathbf{F}}$ denotes the error in the quantization.

The BS assumes that the quantization error matrix $\tilde{\mathbf{F}}$ has $M \times L$ independent identically Gaussian distributed (i.i.d.) elements with zero mean and a variance of $\sigma_E^2/M$. $\sigma_E^2$ is the quantization error variance associated with each quantized vector $\hat{\mathbf{f}}_i$. $\tilde{\mathbf{F}}$ is assumed to be independent of $\mathbf{x}$, $\mathbf{n}$ and $\hat{\mathbf{F}}$. The details of the exact value of $\sigma_E^2$ is described in Appendix A.

The BS designs the linear precoder based on the quantized feedback channels. We at first solve the problem in the virtual uplink and then transfer the solution to downlink using uplink downlink duality. It should be noted here that the linear precoder design with the presence of channel uncertainty model presented in (3.40) was at first solved in [13, 50]. However, we solved it independently in our work. Therefore, we include it in the thesis.

Using our previous equation for the uplink data,

$$x_i^{UL} = \sum_{j=1}^{L} \mathbf{u}_i^H \mathbf{f}_j \sqrt{q_j} x_j + \mathbf{u}_i^H \mathbf{n} \tag{3.41}$$

At the presence of full channel knowledge, MSE of the i$^{th}$ stream takes the following form,

$$E_i^{UL} = \mathbf{u}_i^H \mathbf{F} Q \mathbf{F}^H \mathbf{u}_i + \sigma^2 \mathbf{u}_i^H \mathbf{u}_i + 1 - \mathbf{u}_i^H \mathbf{f}_i \sqrt{q_i} - \sqrt{q_i} \mathbf{f}_i^H \mathbf{u}_i \tag{3.42}$$

However, since we only have the quantized effective channel information $\widehat{\mathbf{F}}$, $E_i^{UL}$ is found as follows,

$$\begin{aligned} E_i^{UL} &= E\left[\mathbf{u}_i^H \left(\hat{\mathbf{F}} + \tilde{\mathbf{F}}\right) \mathbf{Q} \left(\hat{\mathbf{F}}^H + \tilde{\mathbf{F}}^H\right) \mathbf{u}_i | \hat{\mathbf{F}}\right] + E\left[\sigma^2 \mathbf{u}_i^H \mathbf{u}_i + 1 - \mathbf{u}_i^H \left(\hat{\mathbf{f}}_i + \tilde{\mathbf{f}}_i\right) \sqrt{q_i} | \hat{\mathbf{f}}\right] \\ &\quad + E\left[-\sqrt{q_i} \left(\hat{\mathbf{f}}_i^H + \tilde{\mathbf{f}}_i^H\right) \mathbf{u}_i | \hat{\mathbf{F}}\right] \qquad (3.43) \\ &= \mathbf{u}_i^H \hat{\mathbf{F}} \mathbf{Q} \hat{\mathbf{F}}^H \mathbf{u}_i + \sigma^2 \mathbf{u}_i^H \mathbf{u}_i + 1 - \mathbf{u}_i^H \hat{\mathbf{f}}_i \sqrt{q_i} - \sqrt{q_i} \hat{\mathbf{f}}_i \mathbf{u}_i + E\left[\mathbf{u}_i^H \tilde{\mathbf{F}} \mathbf{Q} \tilde{\mathbf{F}}^H \mathbf{u}_i | \hat{\mathbf{F}}\right] \qquad (3.44) \end{aligned}$$

The last equation follows since each of the $\tilde{\mathbf{f}}_i$ vectors are zero mean and uncorrelated with each other and also with the $\hat{\mathbf{f}}$ vectors. Now,

$$\begin{aligned} E_{\tilde{\mathbf{F}}}\left[\mathbf{u}_i^H \tilde{\mathbf{F}} \mathbf{Q} \tilde{\mathbf{F}}^H \mathbf{u}_i | \hat{\mathbf{F}}\right] &= E_{\tilde{\mathbf{F}}}\left[E_{\hat{\mathbf{F}}}\left[\mathbf{u}_i^H \tilde{\mathbf{F}} \mathbf{Q} \tilde{\mathbf{F}}^H \mathbf{u}_i | \hat{\mathbf{F}}, \tilde{\mathbf{F}}\right]\right] \\ &= \frac{1}{M}\left(q_1 \sigma_{E_1}^2 + ... + q_L \sigma_{E_L}^2\right) \mathbf{u}_i^H \mathbf{u}_i \\ &= \frac{1}{M}(q_1 + ... + q_L) \sigma_E^2 \mathbf{u}_i^H \mathbf{u}_i \end{aligned} \tag{3.45}$$

In (3.45), we used the fact that the complex scalar elements of $\tilde{\mathbf{f}}_i$ vectors are i.i.d with zero mean and variance $\frac{\sigma_E^2}{M}$. This holds true since the $\mathbf{f}$ vectors are $M \times 1$ dimensional column vectors and $||\tilde{\mathbf{f}}_i||^2 = \sigma_E^2$ by assumption. Equation (3.45) holds when we assume $\sigma_{E_i}^2$ to be same for different data streams. So,

$$E_i^{UL} = \mathbf{u}_i^H \hat{\mathbf{F}} \mathbf{Q} \hat{\mathbf{F}}^H \mathbf{u}_i + \sigma^2 \mathbf{u}_i^H \mathbf{u}_i + 1 - \mathbf{u}_i^H \hat{\mathbf{f}}_i \sqrt{q_i} - \sqrt{q_i} \hat{\mathbf{f}}_i^H \mathbf{u}_i + \frac{\sigma_E^2}{M}(q_1 + .. + q_L) \mathbf{u}_i^H \mathbf{u}_i \tag{3.46}$$

Differentiating (3.46) with respect to $\mathbf{u}_i^H$ and setting the result to zero, the optimum uplink MMSE filter $\mathbf{u}_i$ is,

$$\mathbf{u}_i^{MMSE} = \mathbf{J}^{-1} \hat{\mathbf{f}}_i \sqrt{q_i} \tag{3.47}$$

Where,

$$\mathbf{J} = \hat{\mathbf{F}}\mathbf{Q}\hat{\mathbf{F}}^H + \sigma^2\mathbf{I}_M + \frac{\sigma_E^2}{M}\left(q_1 + .. + q_L\right)\mathbf{I}_M \tag{3.48}$$

Using $\mathbf{u}_i^{MMSE}$ and $\mathbf{J}$ in (3.46), the MMSE error covariance of data stream $i$ in the uplink is,

$$E_i^{UL,MMSE} = 1 - \sqrt{q}_i\hat{\mathbf{f}}_i^H\mathbf{J}^{-1}\hat{\mathbf{f}}_i\sqrt{q}_i \tag{3.49}$$

The SMSE of the whole system is, therefore

$$SMSE^{UL} = \sum_{i=1}^{L} tr\left[E_i^{UL,MMSE}\right] \tag{3.50}$$

$$= \sum_{i=1}^{L} 1 - \sum_{i=1}^{L} tr\left[\sqrt{q}_i\hat{\mathbf{f}}_i^H\mathbf{J}^{-1}\hat{\mathbf{f}}_i\sqrt{q}_i\right] \tag{3.51}$$

$$= L - tr\left[\mathbf{Q}\hat{\mathbf{F}}^H\left(\hat{\mathbf{F}}\mathbf{Q}\hat{\mathbf{F}}^H + \left(\sigma^2 + \frac{\sigma_E^2\sum_{i=1}^{L} q_i}{M}\right)\mathbf{I}_M\right)^{-1}\hat{\mathbf{F}}\right] \tag{3.52}$$

$$= L - tr\left[\hat{\mathbf{F}}\mathbf{Q}\hat{\mathbf{F}}^H\left(\hat{\mathbf{F}}\mathbf{Q}\hat{\mathbf{F}}^H + \left(\sigma^2 + \frac{\sigma_E^2\sum_{i=1}^{L} q_i}{M}\right)\mathbf{I}_M\right)^{-1}\right] \tag{3.53}$$

$$= L - tr\left[\left(\mathbf{J} - \left(\sigma^2 + \frac{\sigma_E^2\sum_{i=1}^{L} q_i}{M}\right)\mathbf{I}_M\right)\mathbf{J}^{-1}\right] \tag{3.54}$$

$$= L - M + \left(\sigma^2 + \frac{\sigma_E^2}{M}\sum_{i=1}^{L} q_i\right) tr\left[\mathbf{J}^{-1}\right] \tag{3.55}$$

Minimizing the SMSE is therefore equivalent to minimizing $\left(\sigma^2 + \frac{\sigma_E^2}{M}\sum_{i=1}^{L} q_i\right) tr\left[\mathbf{J}^{-1}\right]$. Once $\hat{\mathbf{F}}$ is designed, the SMSE expression is a function of uplink power allocation $\mathbf{Q}$.

The optimization problem for power allocation is,

$$\mathbf{Q}^{opt} = \arg\min_{\mathbf{Q}}\left(\sigma^2 + \frac{q_1 + ... + q_L}{M}\sigma_E^2\right) tr(\mathbf{J}^{-1}) \tag{3.56}$$

$$subject\,to\,: tr\left[\mathbf{Q}\right] \leq P_{max}, q_i \geq 0 \forall i \in [1, L]$$

Ding [13] shows that SMSE remains a nonincreasing function of SNR if all available power is used i.e. $tr\left(\mathbf{Q}\right) = \sum q_i = P_{max}$. Therefore, the optimization problem remains convex since, the term $\left(\sigma^2 + \frac{q_1+...+q_L}{M}\sigma_E^2\right)$ becomes a constant. The convexity with respect to $\mathbf{J}$ is proved in [7]. Using [57], the optimal $\mathbf{p} = \mathbf{q}$.

## 3.6 Receiver design for data processing

As mentioned earlier, MESC is for quantization purposes only. The base station determines $\mathbf{p}$ and $\mathbf{U}$ based on the quantized $\widehat{\mathbf{F}}$. However, for mutually nonorthogonal reported channels and a finite number of users, using MMSE receivers for data processing provide better results than MESC receivers [58]. Using the symbol policy of (2.2), for the data,

$$\mathbf{v}_{k_j} = \left(\mathbf{H}_k^H \mathbf{U}\mathbf{P}\mathbf{U}^H \mathbf{H}_k + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{H}_k^H \mathbf{u}_{k_j} \sqrt{p}_{k_j}, \tag{3.57}$$

which can be normalized to make $||\mathbf{v}_{k_j}|| = 1$. Note that the MMSE receiver cannot be implemented at the time of channel quantization since the precoder matrix $\mathbf{U}$ was not designed at that time.

The implementation of the decoder mentioned in (3.57) requires infinite training symbols. Therefore, from a practical point of view, the BS either sends a finite number of dedicated symbols [37] or uses limited feedforward [8] to convey the post-processing information to the receivers. However, in our simulations, we restrict ourselves to the case where the users can estimate the effective channels of their data streams.

Note that, our overall algorithm is sub-optimal because $\mathbf{U}$ and $\mathbf{p}$ are designed using MESC, not MMSE. This is the price paid for the feedback to be independent across users.

## 3.7 Analysis and Discussion

### 3.7.1 Different Channel Model Assumptions at the Base Station and at the Receiver End

The channel model, representing the relation between the original channel, quantized channel and error in channel is as follows:

$$\mathbf{f}_i = ||\mathbf{f}_i|| \left( \left( \hat{\mathbf{f}}_i^H \overline{\mathbf{f}}_i \right) \hat{\mathbf{f}}_i + \tilde{\mathbf{f}}_i \right) \tag{3.58}$$

Since the receiver knows the channel exactly, it uses the original channel model. However, we use the following channel model at the base station,

$$\mathbf{f}_i = \hat{\mathbf{f}}_i + \tilde{\mathbf{f}}_i \tag{3.59}$$

$\tilde{\mathbf{f}}_i$ is an unknown vector at the BS.

There are two difference between (3.58) and (3.59). They are as follows:

1. Our simulation results show that expending all the available bits in the direction of the channel provide better performance, in terms of bit error rate, in the proposed system model that use PSK based modulation. Therefore, we only provide unit norm shape feedback in this part of our work. So the BS is unaware of the channel norm $||\mathbf{f}_i||_2$. In chapter 4, we adopt a more theoretical approach and optimally allocate bits in the norm and shape of the channel.

2. To minimize the feedback overhead, the receiver does not expend any bit in quantizing $\left(\hat{\mathbf{f}}_i^H \overline{\mathbf{f}}_i\right)$. Therefore, there is a phase shift of $\left(\hat{\mathbf{f}}_i^H \overline{\mathbf{f}}_i\right)$ between the original channel model of (3.58) and assumed channel model of (3.59). However, since we propose using a *MMSE decoder while actually receiving data*, system performance automatically compensates for this phase shift.

### 3.7.2 Relation to the Existing Algorithms

As the proposed receive combining technique maximizes the expected SINR of the data streams at the user end, it is equivalent to the MESC algorithm in the case of one data stream per user of [58] which was designed for the zero forcing (ZF) precoder. To illustrate this, let $L_k = 1$. Since intra-user interference is not present, all the quantized effective channels in $\widehat{\mathbf{F}}$ are assumed to be mutually orthogonal. Using this in (3.20) we get,

$$\begin{aligned}\mathbf{u}_{k_j} &= \frac{1}{\sigma^2 + \sigma_E^2 P_{max}} \hat{\mathbf{f}}_{k_j} \sqrt{q}_{k_j} - \frac{1}{\left(\sigma^2 + \sigma_E^2 P_{max}\right)^2} \hat{\mathbf{F}} \left(\mathbf{Q}^{-1} + \frac{1}{\sigma^2 + \sigma_E^2 P_{max}} \mathbf{I}\right)^{-1} [1, 0, \cdots, 0]^T \sqrt{q}_{k_j} \\ &\propto \hat{\mathbf{f}}_{k_j}, \end{aligned} \tag{3.60}$$

(3.60) follows since $\left(\mathbf{Q}^{-1}+\frac{1}{\sigma^2+\sigma_E^2 P_{max}}\mathbf{I}\right)^{-1}$ is a diagonal matrix. Since $||\mathbf{u}_{k_j}||=1$, $\mathbf{u}_{k_j}=\hat{\mathbf{f}}_{k_j}$ in this scenario. Using this in (3.21) we find,

$$SINR_{k_j}^{DL}=\frac{\mathbf{v}_{k_j}^H\left(\frac{P}{L}\mathbf{H}_k^H\hat{\mathbf{f}}_{k_j}\hat{\mathbf{f}}_{k_j}^H\mathbf{H}_k\right)\mathbf{v}_{k_j}}{\sigma^2+\mathbf{v}_{k_j}^H\left(\frac{P}{L}\mathbf{H}_k^H\left(\frac{L-1}{M-1}\left(\mathbf{I}-\hat{\mathbf{f}}_{k_j}\hat{\mathbf{f}}_{k_j}^H\right)\right)\mathbf{H}_k\right)\mathbf{v}_{k_j}} \tag{3.61}$$

This is the expression obtained in [58] as the MESC combiner with noise variance $\sigma^2=1$. [58] has shown that this algorithm takes the form of MET combining at low SNR and QBC at high SNR. Thus MESC combining of [58] considers signal power and inter-user interference while choosing the code vector. Since we are considering multiple data streams to each user, our proposed SINR expression in (3.21) considers signal power, inter-user and intra-user interference altogether. Thus our proposed algorithm is a generalized form for MESC combining with multiple data streams.

### 3.7.3 SMSE Analysis

In the absence of quantization error, the SMSE of the precoder with perfect CSI is [32]:

$$\begin{aligned}SMSE&=L-M+\sigma^2 tr\left[\left(\mathbf{F}\mathbf{Q}\mathbf{F}^H+\sigma^2\mathbf{I}_M\right)^{-1}\right]\\&=L-M+tr\left[\left(\frac{P_{max}}{L\sigma^2}\mathbf{F}\mathbf{F}^H+\mathbf{I}_M\right)^{-1}\right]\end{aligned} \tag{3.62}$$

In (3.62), we assumed $\mathbf{Q}=(P_{max}/L)\,\mathbf{I}_L$ i.e., equal power allocation for simplicity of the analysis. $tr\left(\frac{P_{max}}{L\sigma^2}\mathbf{F}\mathbf{F}^H+\mathbf{I}_M\right)^{-1}$ is a decreasing function of SNR and hence SMSE decreases with SNR. However, with quantization error, if the original precoder [32] is used,

$$SMSE=\sum_{i=1}^{L}\left(1-q_i\widehat{\mathbf{f}}_i^H\mathbf{J}^{-1}\widehat{\mathbf{f}}_i+\frac{\sigma_E^2}{M}P_{max}q_i\widehat{\mathbf{f}}_i^H\mathbf{J}^{-2}\widehat{\mathbf{f}}_i\right) \tag{3.63}$$

where $\mathbf{J}=\widehat{\mathbf{F}}\mathbf{Q}\widehat{\mathbf{F}}^H+\sigma^2\mathbf{I}_M$. Both $q_i\widehat{\mathbf{f}}_i^H\mathbf{J}^{-1}\widehat{\mathbf{f}}_i$ and $\frac{\sigma_E^2}{M}P_{max}q_i\widehat{\mathbf{f}}_i^H\mathbf{J}^{-2}\widehat{\mathbf{f}}_i$ increase with SNR. Since the former term is a linear-over-affine function and the latter is a quadratic-over-quadratic function of $P_{max}$, at high SNR the latter term dominates and SMSE increases with SNR.

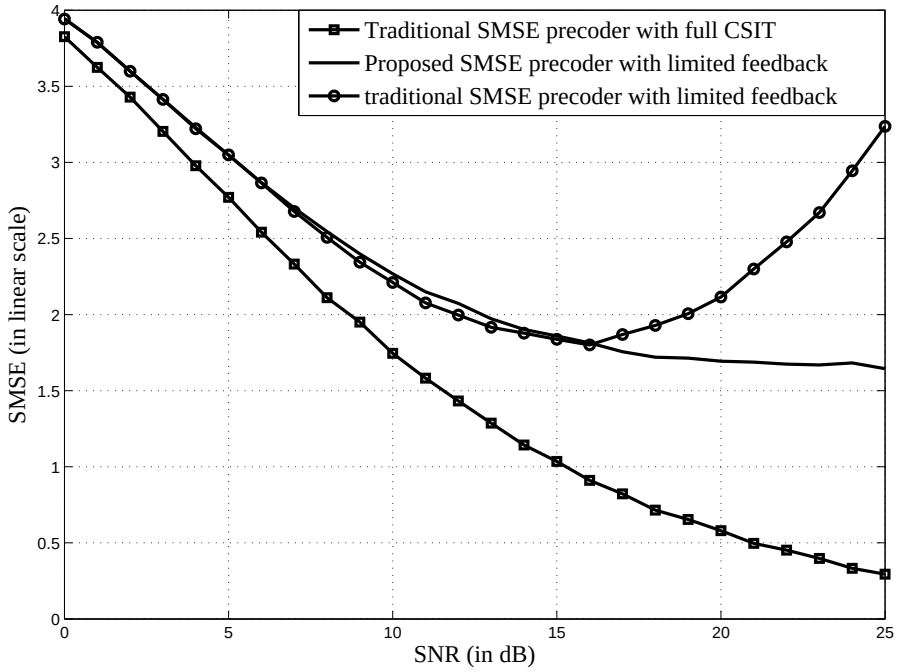



Figure 3.2: Comparing the SMSE of the traditional and the proposed precoder, ($M = 5$, $K = 5$, $N_k = 1$, $L_k = 1\ \forall\, k$, $B = 10$, QPSK)

Some recent works on SMSE based precoder design with channel uncertainty [11] observed this effect but did not analyze it.

In our proposed algorithm,

$$SMSE = L - M + \left(\sigma^2 + \frac{\sigma_E^2}{M} P_{max}\right) tr\left[\left(\hat{\mathbf{F}}\mathbf{Q}\hat{\mathbf{F}}^H + \left(\sigma^2 + \frac{\sigma_E^2}{M} P_{max}\right)\mathbf{I}_M\right)^{-1}\right] \tag{3.64}$$

$$= L - M + tr\left(\frac{P_{max}}{L\left(\sigma^2 + \frac{\sigma_E^2}{M} P_{max}\right)}\hat{\mathbf{F}}\hat{\mathbf{F}}^H + \mathbf{I}_M\right)^{-1} \tag{3.65}$$

In (3.65), we again assumed equal power allocation for analysis simplicity. $\frac{P_{max}}{L\left(\sigma^2 + \frac{\sigma_E^2}{M} P_{max}\right)}$ is a nonincreasing function of $P_{max}$. Thus the proposed precoder makes sure that SMSE does not increase with SNR in the high SNR region. Fig. 3.2 illustrates all these effects. Since, the increase in SMSE is most apparent in MU-MISO systems, the simulations use a MU-MISO system with independent channel realizations where $M = 5$, $L_k = 1\ \forall k$ and

$B = 10$ bits per data stream. We use uncoded QPSK for data transfer. The proposed algorithm clearly stabilizes the SMSE at high SNR.

## 3.8 Numerical Simulations

In this section we compare our proposed scheme with the leading feedback schemes in the literature. Since our proposed algorithm uses an MMSE based receiver at the data transmission phase, we use an MMSE receiver to simulate the other existing algorithms, too. This preserves the fairness of the comparisons since the performance of the system always improves with an MMSE receiver for mutually non-orthogonal channels [58]. Unless specified, all transmissions use uncoded QPSK.

As mentioned before, our proposed transceiver for MU - MIMO systems can be readily generalized to MU - MISO systems. In Fig. 3.3, we compare the performance of the proposed algorithm to some of the available precoders in a limited feedback MU - MISO system. The system uses $M = 4$, $K = 4$, $L_k = 1\ \forall k$ and $B = 10$. The proposed algorithm performs better than the MMSE precoder [11] by using MSIP quantization and convexity of the power allocation problem. The traditional SMSE transceiver, that ignores quantization error, performs well at low SNR, but begins to worsen at a SNR of 15dB. Thus the proposed transceiver improves over the state-of-the-art in MU MISO precoders based on limited feedback.

To the best of our knowledge, coordinated beamforming [9] is one of the very few existing linear transceivers that avoids the dimensionality constraint in the MU MIMO with multiple data stream per user scenario. In Fig. 3.4 we compare the proposed algorithm with coordinated beamforming. Here $M = 4$, $K = 2$, $N_k = 4$, $L_k = 1\ \forall k$ and $B = 15$. Since coordinated beamforming implements joint transceiver design, it performs better than the proposed algorithm with full CSIT. However, coordinated beamforming needs at least $(M^2 - 1)$ bits for the feedback of $\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^H/||\hat{\mathbf{H}}||_F^2\right)$. To create Fig. 3.4 we used

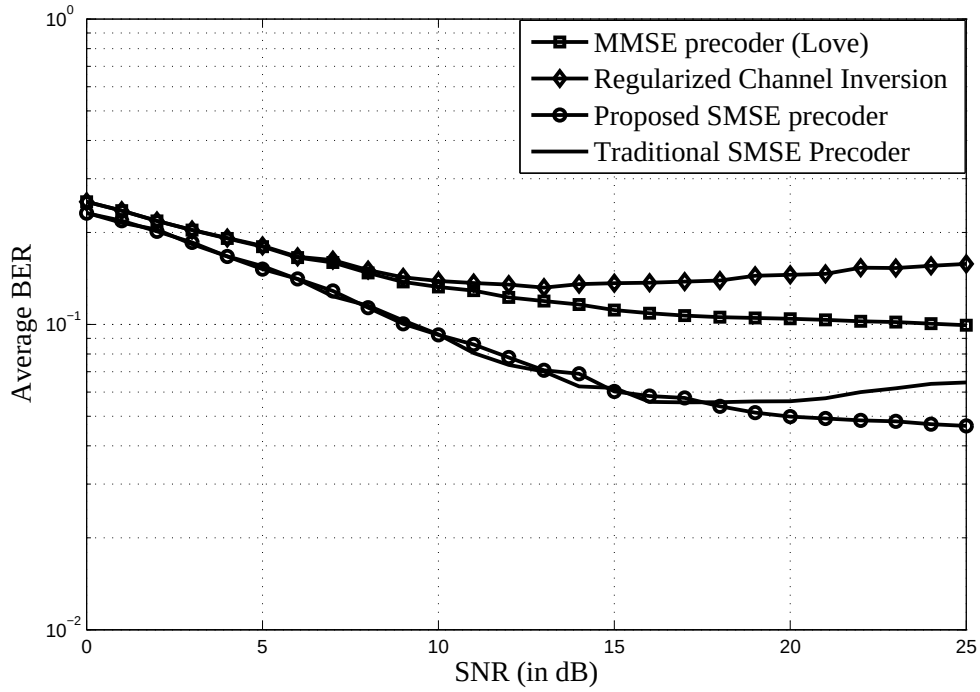



Figure 3.3: Comparison with available MU-MISO precoding techniques ($M = 4$, $K = 4$, $L_k = 1\ \forall\, k$, $B = 10$, QPSK)

15 bits feedback overhead per data stream in a MU MIMO system with four transmit antennas. This means only 1 bit is available per unique scalar entry of $\left(\hat{\mathbf{H}}\hat{\mathbf{H}}^H/||\hat{\mathbf{H}}||_F^2\right)$ , introducing large quantization error. The eigen structure of the channel therefore gets mangled at the BS [40], leading to loss of performance. On the other hand, since our proposed algorithm expends 15 bits to quantize the $4\times 1$ vector, the quantization error of the fed back vector always remains less than or equal to $2^{\frac{-B}{M-1}} = 0.03125$. Thus, the proposed algorithm performs very close to its full CSIT curve and outperforms coordinated beamforming [9] with limited feedback.

In Fig. 3.5 we compare our proposed scheme with other VQ combining limited feedback MU MIMO transceivers. In this example, $M = 4$, $K = 2$, $N_1 = N_2 = 2$, $N_3 = 3$, $L_k = 1\ \forall k$ and $B = 15$. Since to the best of our knowledge, existing VQ combining MU

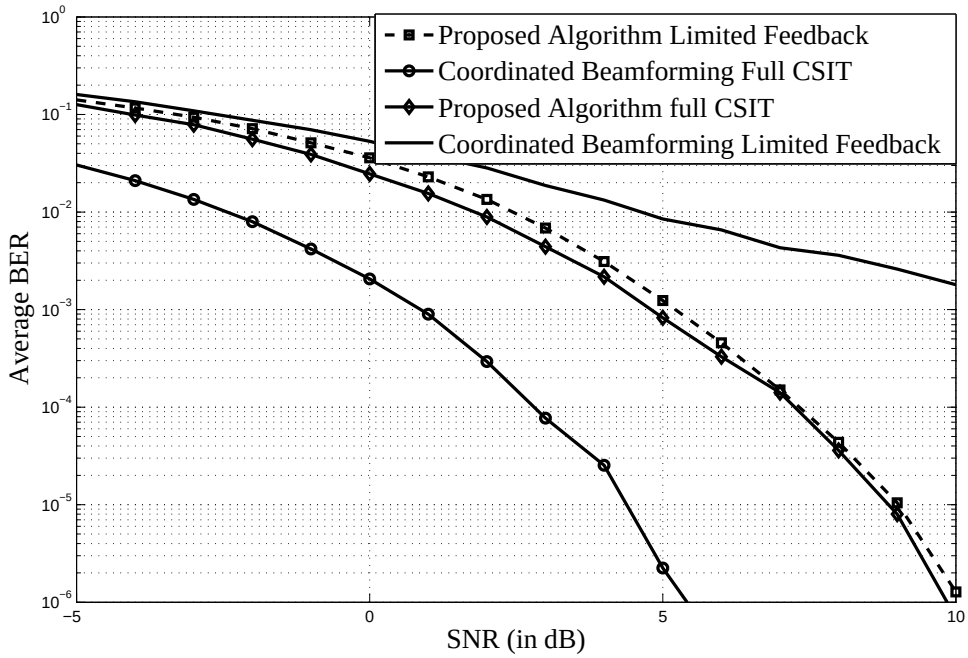



Figure 3.4: Comparison with the coordinated beamforming ($M = 4$, $K = 2$, $N_k = 4$, $L_k = 1\ \forall\, k$, $B = 15$, QPSK)

MIMO schemes have not dealt with multiple data streams per user, we stick with one data stream per user in this comparison. The proposed scheme outperforms the QBC [30] and MET [5] approaches due to the use of SMSE precoder, adaptive receive combining and optimal power allocation. Although our algorithm outperforms Boccardi's MESC [58] up to 20 dB, [58] seems to converge at a lower error floor than the proposed algorithm. This happens because in our proposed algorithm the actual quantization error variance is not known at the BS. Due to the adaptive quantization policy of the proposed algorithm, the quantization error variance changes from low to high SNR; since we only quantize the direction of the effective channels, the norm of the quantization error is not available at the BS. The quantization error in MESC case [58] also changes from low to high SNR but the BS does not need this knowledge due to the use of a ZF precoder.

Our proposed transceiver adds to the literature by allowing multiple data streams per

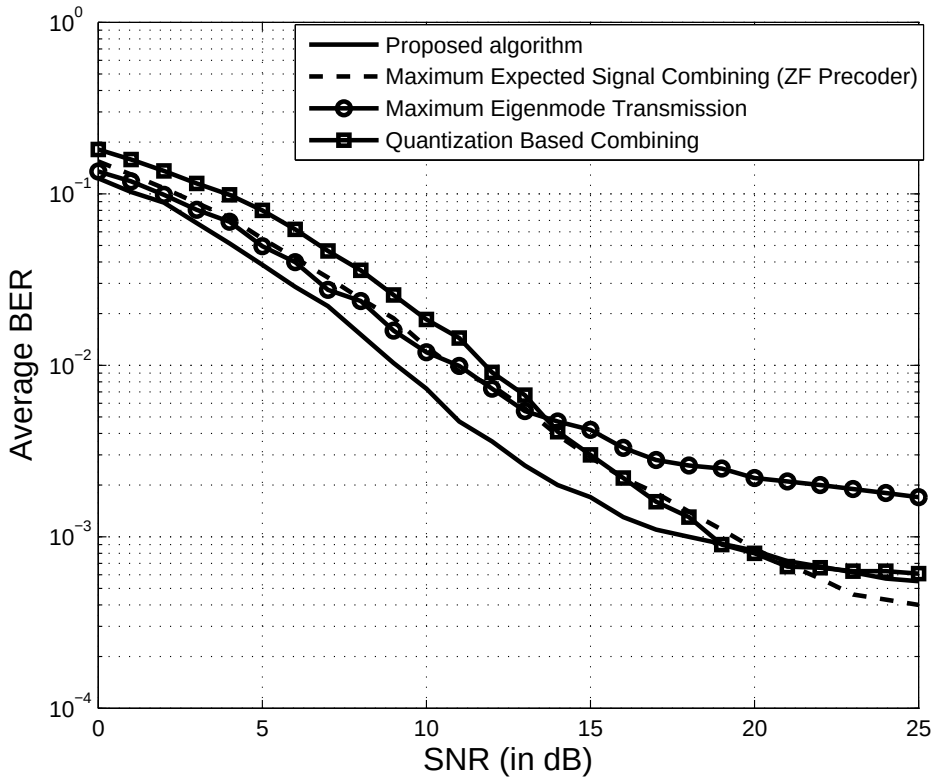



Figure 3.5: Comparison with available MU-MIMO VQ precoding techniques ($M = 4$, $N_1 = N_2 = 2$, $K = 3$, $N_3 = 3$, $L_k = 1$, $\forall\, k$, $B = 15$, QPSK)

user. Fig. 3.6 shows the comparison of the transceiver's performance to other possible methods to transmit multiple data streams per user. In Fig. 3.6, Eigen Based Combining (EBC) projects the MIMO channel to its dominant eigenvectors to create effective MISO channels [25] and QBC chooses the set of codevectors that will generate least quantization error as effective MISO channels [30]. The proposed transceiver approaches EBC at low SNR and QBC at high SNR. Thus the proposed algorithm retains the advantages of both EBC and QBC by providing a trade-off between signal power, intra-user and inter-user interference.

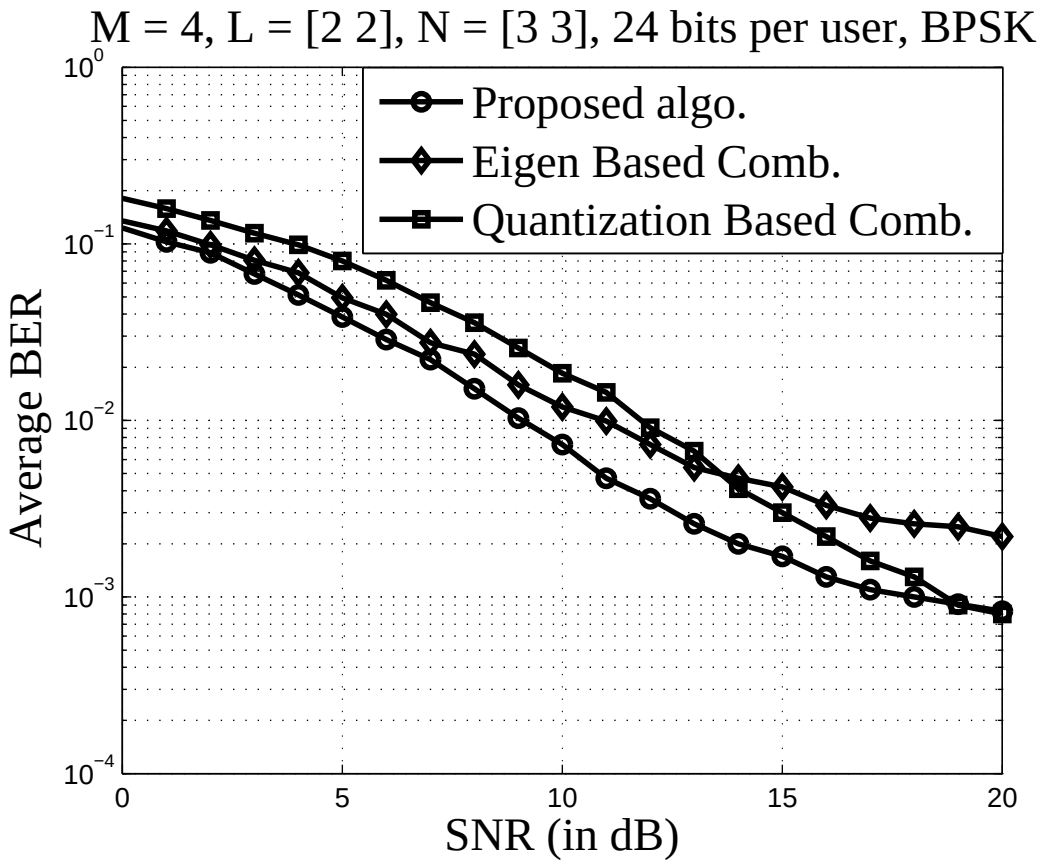



Figure 3.6: Different receive combining techniques with multiple data streams per user ($M = 4$, $N_k = 3$, $L_k = 2$, $\forall\, k$, $B = 12$, BPSK)

# Chapter 4

# Optimal Bit Allocation across Gain and Shape Feedback

We used two heuristic approaches in the previous chapter. These approaches were described in section 3.7.1. We again mention it here:

1. We used all the available bits to quantize the shape of the individual user's channel, assuming that the channel gains do not play a significant role.

2. We used MSIP based feedback and SMSE based precoder in our overall system design. Although the presence of phase shift between the original and quantized channel is an inherent property of the MSIP based feedback, the overall system performance, in terms of BER, would not be affected since we used MMSE decoder as the receiver.

In this chapter, we approach the quantization problem from a more theoretical perspective. Our objective here is the investigation of the effect of optimal bit allocation across gain and shape in the performance of a multiuser system. Therefore, we use a product based codebook and quantize the gain and shape of the channel separately. We provide optimal bit allocation across gain and shape feedback to minimize the overall SMSE of the system.

The difference between (3.58) and (3.59) arises due to the use of chordal distance.

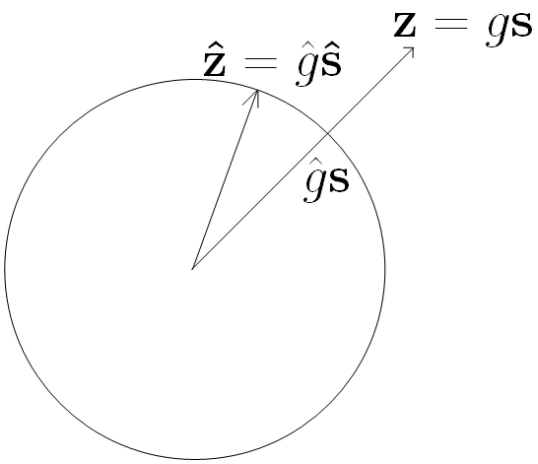



Figure 4.1: Separate gain and shape quantization using product quantization

Therefore, we use Euclidean distance in this part of our work. To the best of our knowledge, this is the only work that provides optimal bit allocation based on euclidean distance among the norm and shape of a complex channel.

## 4.1 Problem Statement

We use a product based codebook in this work. Therefore, we have separate codebooks to quantize the gain and shape of the channel. Let us assume that we expend $B_s$ and $B_g$ bits to quantize the gain and shape of the channel respectively. Now, $B = B_s + B_g$ where $B$ is the total feedback overhead per data stream. Therefore, $N_s = 2^{B_s}, N_g = 2^{B_g}$. Here, $N_s$ and $N_g$ are the total number of shape and gain codevectors respectively.

From the previous chapter,

$$SMSE = L - M + \left(\sigma^2 + \frac{\sigma_E^2}{M}\sum_{i=1}^{L} q_i\right) tr\left[\mathbf{J}^{-1}\right] \tag{4.1}$$

The previous chapter dealt with the design of $\mathbf{P}, \mathbf{U}, \mathbf{V}$ to minimize (4.1) for a given $\sigma_E^2$. In this chapter, we focus on minimizing $\sigma_E^2$ using optimal bit allocation.

*Problem Statement:*

Let $\mathbf{z} \in \mathbb{C}^M$ represent the original channel vector to be quantized. Comparing with

the notations used in the previous chapters, $\mathbf{z}$ will take the form of $\mathbf{h}$ in the MISO case. On the other hand, $\mathbf{z}$ takes the form of $\mathbf{f}$ ($\mathbf{f} = \mathbf{Hv}$) i.e., it represents the effective vector downlink channel of the original MIMO channel $\mathbf{H}$. As specified in section 2.5, due to the lack of knowledge of channel norm in MESC combining, we used eigen-based combining in this part of our work. Therefore, $\mathbf{z}$ represents the product of singular value and singular vector in the MIMO case.

Let $\hat{\mathbf{z}}$ represent the quantized channel. Let, $\mathcal{C}$ be the quantization codebook. The original problem statement is as follows:

$$min_{Bs,Bg} E\left[||\mathbf{z} - \hat{\mathbf{z}}||^2\right] \tag{4.2}$$
$$subject\,to: B_s + B_g = B\,,\ \hat{\mathbf{z}} \in \mathcal{C}$$

Fig. 4.1 represents the product codebook operation based on separate gain and shape quantization. Let, $\mathbf{z} = g\mathbf{s}$. Here, $g$ and $\mathbf{s}$ denote the gain and shape of the channel respectively, i.e., $g$ is positive and $||\mathbf{s}||_2 = 1$. Due to the use of separate gain and shape quantization and a product based codebook, the BS finds $\hat{\mathbf{z}}$ as, $\hat{\mathbf{z}} = \hat{g}\hat{\mathbf{s}}$. Here, $\hat{g}$ and $\hat{\mathbf{s}}$ denote the quantized gain and shape respectively.

If we consider a MIMO system, these will indicate the singular values and the directions of the singular vectors (i.e., the $\mathbf{f}$ vectors) of the MIMO channel respectively. If we consider a MISO system, $g$ and $\mathbf{s}$ will indicate the norm and direction of the original channels (i.e., the $\mathbf{h}$ vectors) respectively.

The Lloyd-Max algorithm [39] is the optimal solution to find the codebook for the gain of the vector. Simulation results show that numerically achieved codebook based on the faster K-means algorithm [42] also provides almost similar performance. We use K-means algorithm in our work.

The Euclidean distance based optimal codebook of unit norm codevectors is not yet known. Therefore, we use random VQ to find the shape codebook of the channel i.e., the unit norm quantized shape codevectors of the codebook are randomly and independently

distributed in $\mathbb{C}^M$.

Let $\mathcal{C}_g$ and $\mathcal{C}_s$ represent the gain and shape codebook respectively. Note that, the size and form of $\mathcal{C}_g$ and $\mathcal{C}_s$ will vary with respect to $B_s$ and $B_g$.

Now, the optimal bit allocation problem takes the following form,

$$min_{Bs,Bg} E\left[||\mathbf{z} - \hat{g}\hat{\mathbf{s}}||^2\right] \tag{4.3}$$
$$subject\,to: B_s + B_g = B\,,\ \hat{g} \in \mathcal{C}_g\,,\ \hat{\mathbf{s}} \in \mathcal{C}_s$$

Hamkins et al. [20] has shown that, for high rate quantization, the quantization distortion takes the following form,

$$E\left[||\mathbf{z} - \hat{g}\hat{\mathbf{s}}||^2\right] \approx E\left[(g-\hat{g})^2\right] + E\left[g^2\right] E\left[||\mathbf{s} - \hat{\mathbf{s}}||^2\right] \tag{4.4}$$
$$\approx D_g + E\left[g^2\right] D_s \tag{4.5}$$

Here, $E[g^2]$ denotes the variance of the gain. $D_g = E[(g-\hat{g})^2]$ denotes the distortion due to gain quantization. $D_s = E[||\mathbf{s} - \hat{\mathbf{s}}||^2]$ represents the distortion due to unit norm shape quantization. Since $D_g$ and $D_s$ are independent of each other in (4.5), the optimal bit allocation problem can be solved using the following three steps:

1. Find $D_g$, gain distortion, for a given $B_g$.
2. Find $D_s$, shape distortion, for a given $B_s$.
3. Provide optimal bit allocation to minimize the overall distortion, i.e., $min_{B_s,B_g} E[g^2]\, D_s + D_g$.

We will provide descriptions of those three steps in the next few sections.

## 4.2 Distortion due to gain quantization

Distortion due to gain quantization is given through the following equation,

$$D_g = E\left[(g-\hat{g})^2\right] \tag{4.6}$$
$$= \int_0^\infty \left(r - \hat{g}(r)\right)^2 f_g(r) dr \tag{4.7}$$

Here, $r$ is the random variable representing gain. $f_g(r)$ is the probability density function of gain. Using Bennett's integral ( [15], page-186), the gain distortion of (4.7) takes the following form,

$$D_g = \frac{1}{12N_g^2}||f_g(r)||_{\frac{1}{3}} \tag{4.8}$$

Here, $||f_g(r)||_{\frac{1}{3}}$ denotes $\left(\int_0^\infty |f_g(r)|^{\frac{1}{3}}dr\right)^3$. The distortion of the norm of a MISO channel due to quantization has been calculated by Hamkins et al. in [20]. Following Hamkins' derivation, we seek to find the analytical expression of the quantization distortion of the eigenvalues of MIMO channel.

*Lemma 1:* Using the probability distribution of the dominant eigenvalues of Wishart matrix [54] and Jacobian transform [52],

$$||f_g(r)||_{\frac{1}{3}} = \frac{3 \times 3^{L(e)}\beta}{4(L(e)-1)!}\Gamma^3\left(\frac{L(e)+1}{3}\right) \tag{4.9}$$

Where, $L(e) = (M-e)(N-e)$. $M$ and $N$ are the number of transmit and receive antennas respectively. $e$ denotes the order of the eigenvalue where 0 represents the most dominant one, 1 denotes the 2nd most dominant one and so on. Here, $\beta = \frac{\tilde{\lambda_e}}{L(e)}$ where $\tilde{\lambda_e}$ is the mean of the e$^{th}$ eigenvalue.

*Proof:* See Appendix B.

Using (4.8) and (4.9), the gain distortion at high resolution can be expressed as,

$$D_g = \frac{1}{12N_g^2}||f_g(r)||_{\frac{1}{3}} \tag{4.10}$$

$$= \frac{1}{16N_g^2}\frac{3^{L(e)}\beta}{(L(e)-1)!}\Gamma^3\left(\frac{L(e)+1}{3}\right) \tag{4.11}$$

$$= C_g 2^{-2B_g} \tag{4.12}$$

Here, $C_g = \frac{1}{16}\frac{3^{L(e)}\beta}{(L(e)-1)!}\Gamma^3\left(\frac{L(e)+1}{3}\right)$ is a constant with respect to $B_g$.

Note that the gain distortion of a complex MISO vector due to quantization was obtained in [20] through the following equation,

$$D_g = C_{g_{MISO}} 2^{-2B_g} \tag{4.13}$$

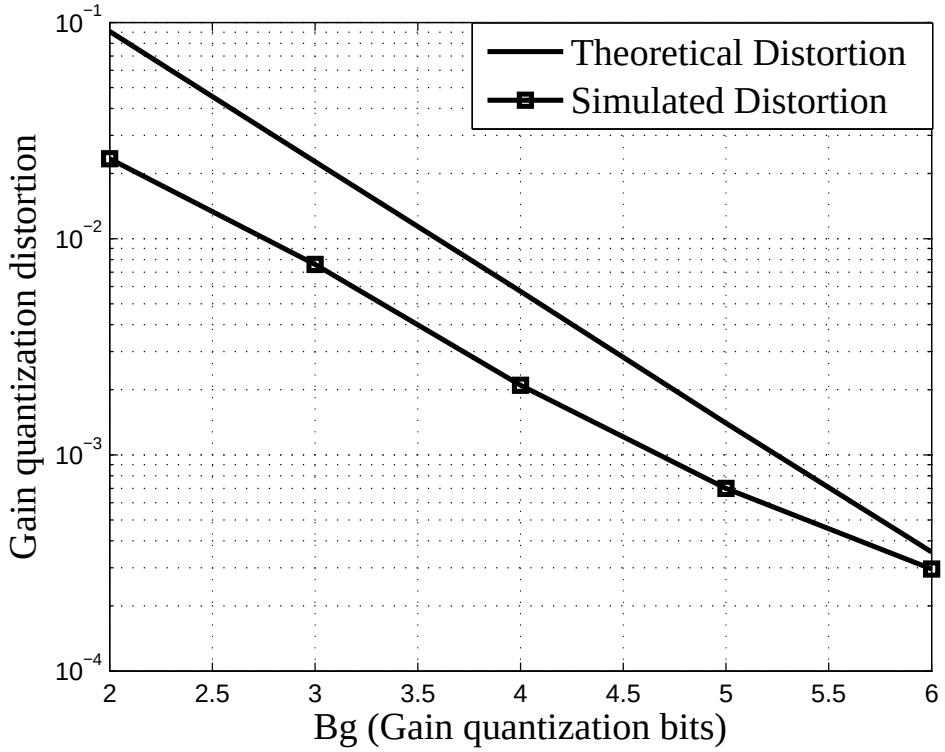



Figure 4.2: Quantization distortion of the dominant singular value of 2x2 MIMO channel

Here, $C_{g_{MISO}} = \frac{3^{k/2}\Gamma\left(\frac{k+2}{6}\right)}{8\Gamma\left(\frac{k}{2}\right)}$ [20]. Note that both (4.12) and (4.13) suggest that gain distortion due to quantization is proportional to $2^{-2B_g}$.

Fig. 4.2 shows the distortion due to gain quantization of the dominant singular value of a $2 \times 2$ MIMO channel. According to the figure, the analytical expression starts to converge with the simulated result as $B_g$ increases. This observation matches with the fact that the Bennett integral in (4.9) holds for high bit quantization.

## 4.3 Distortion due to Shape Quantization

This section focuses on the shape quantization error of a unit norm vector located in $\mathbb{C}^M$.

Here, we are measuring the quantization error of two unit norm vectors in terms of Euclidean distance. The Euclidean distance of two points in a $\mathbb{C}^M$ plane has a one-to-

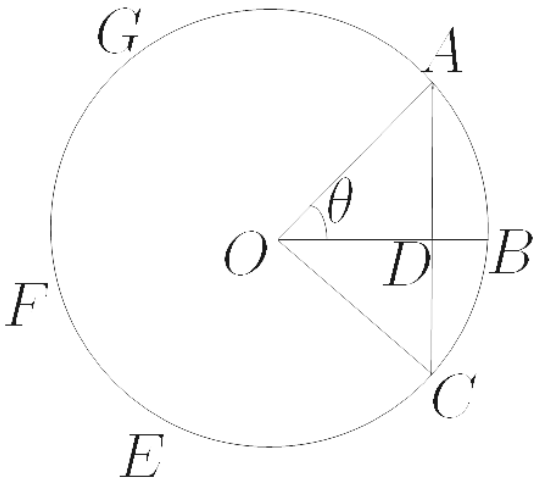



Figure 4.3: Shape quantization block diagram

one correlation with that of two points in a $\mathbb{R}^{2M}$ plane. Therefore, from now on, we will assume that we are dealing with vectors in the $\mathbb{R}^{2M}$ plane.

Now, let $\mathbf{s}$ and $\hat{\mathbf{s}}$ denote the original and quantized channel vectors. Figure 4.3 shows a two dimensional view of the problem that we are trying to solve. Here, OB and OA denote $\mathbf{s}$ and $\hat{\mathbf{s}}$ respectively.

Let us assume that the Euclidean distance between $\mathbf{s}$ and $\hat{\mathbf{s}}$ is $d$ i.e.,

$$d = ||\mathbf{s} - \hat{\mathbf{s}}||_2 \tag{4.14}$$

Define $\mathcal{U}_{2M}$ i.e., $ABCEFG$ as the unit hypersphere in $\mathbb{R}^{2M}$. The surface area of $\mathcal{U}_{2M}$ i.e., $SA\left(ABCEFG\right)$ is given by [33],

$$SA\left(\mathcal{U}_{2M}\right) = 2MC_{2M} \tag{4.15}$$

Where,

$$C_{2M} = \frac{\pi^M}{\Gamma(M+1)} \tag{4.16}$$

Define the spherical caps $\mathcal{D}^{2M}$ i.e., $ABC$ around the channel vector $\mathbf{s}$,

$$\mathcal{D}^{2M} = (\hat{\mathbf{s}} \in \mathcal{U}_{2M} | ||\mathbf{s} - \hat{\mathbf{s}}||_2 \leq d) \tag{4.17}$$

In hypersphere of real dimensions, there is a one to one correlation between the Euclidean and angular distance between two points. In Fig. 4.3, let $\angle AOB = \theta$ be the angular distance between $\mathbf{s}$ and $\hat{\mathbf{s}}$. Since $||OA|| = ||\hat{\mathbf{s}}|| = 1$, $AD = sin(\theta)$ and $OD = cos(\theta)$. Since $||OB|| = ||\mathbf{s}|| = 1$, $BD = 1 - cos(\theta)$. Therefore,

$$AB^2 = AD^2 + BD^2 = sin^2(\theta) + (1 - cos(\theta))^2 = 2 - 2cos(\theta) \tag{4.18}$$

Assuming $b = d^2$,

$$b = 2 - 2cos(\theta) \tag{4.19}$$

$$\theta = cos^{-1}(1 - 0.5b) \tag{4.20}$$

The surface area of the spherical cap $\mathcal{D}^{2M}$ i.e., $SA(ABC)$ is given by [33],

$$SA\left(\mathcal{D}^{2M}\right) = (2M-1)C_{2M-1}\int_0^{\theta} sin^{2M-2}\phi d\phi \tag{4.21}$$

We use random VQ in the proposed problem. The quantization code vectors are uniformly and independently distributed in $\mathbb{C}^M$. So, the quantization code vectors in $\mathbb{R}^{2M}$ are also independent and isotropically distributed. Therefore, if we assume a small sphere of radius $d$ centred on $\mathbf{s}$, the quantized code vectors can lie anywhere in this sphere. This leads to the following result,

$$Pr[||\mathbf{s} - \hat{\mathbf{s}}||^2 \leq b] = \frac{SA(ABC)}{SA(ABCEFG)} \tag{4.22}$$

Using (4.15), (4.16), (4.20) and (4.21) in (4.22) we get,

$$Pr[||\mathbf{s} - \hat{\mathbf{s}}||^2 \leq b] = \frac{(2M-1)C_{2M-1}\int_0^{cos^{-1}(1-0.5b)} sin^{2M-2}\phi d\phi}{2MC_{2M}} \tag{4.23}$$

All the quantized codevectors are randomly chosen. Therefore, the probability that the square of the Euclidean distance between any quantization codevector and the original channel is higher than $b$, is independent of the other. Therefore,

$$Pr[\min_{i\in[1,N_s]} ||\mathbf{s} - \hat{\mathbf{s}}_i||^2 \geq b] = \left(1 - \frac{(2M-1)C_{2M-1}\int_0^{cos^{-1}(1-0.5b)} sin^{2M-2}\phi d\phi}{2MC_{2M}}\right)^N \tag{4.24}$$

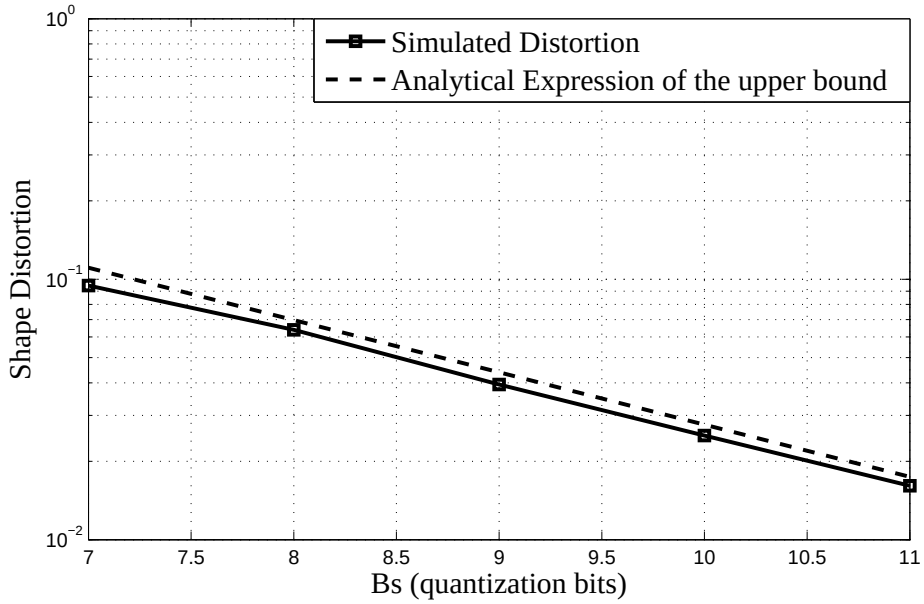



Figure 4.4: Comparison of the simulated distortion with the theoretical upper bound (2x1 complex vector)

Now, expected value of the distortion error variance due to shape quantization is found as follows,

$$E(b) = \int_0^4 Pr[\min_{i \in N} ||\mathbf{s} - \hat{\mathbf{s}}_i||^2 \geq b] db \tag{4.25}$$

The limits of integration in (4.25) follows from the fact that the square of the Euclidean distance between two points in a unit radius complex sphere ranges between 0 and 4.

*Lemma 2*:

$$E(b) \leq C_s 2^{\frac{-2B_s}{2M-1}} \tag{4.26}$$

Here, $C_s = \left( \frac{\pi^{\frac{2M-1}{2}} \Gamma(M)}{2\pi^M \Gamma\left(\frac{2M-1}{2}+1\right)} \right)^{\frac{-2}{2M-1}}$ is a constant with respect to $B_s$.

Proof: See Appendix B.

Eq. (4.26) leads to the following result,

$$E\left(||\mathbf{s} - \hat{\mathbf{s}}||^2\right) < \left( \frac{\pi^{\frac{2M-1}{2}} \Gamma(M)}{2\pi^M \Gamma\left(\frac{2M-1}{2}+1\right)} \right)^{\frac{-2}{2M-1}} 2^{\frac{-2B_s}{2M-1}} \tag{4.27}$$

Fig. 4.4 compares the derived analytical result with simulated distortion. In the simulation, we generate $2^{B_s}$ random unit norm quantization code vector for each shape quantization bit, $B_s$, shown in fig. 4.4. We assume that this set of code vector constitutes the codebook. Thereafter, we generate a random unit norm vector downlink channel, $\mathbf{s}$, and map it to the quantized code vector based on Euclidean distance i.e. we find the quantized code vector of the codebook that has the least Euclidean distance with $\mathbf{s}$. We calculate the squared Euclidean distance between $\mathbf{s}$ and its corresponding quantized code vector. We iterate this process, i.e., generate different unit norm vector downlink channel for 1000 times and find the average squared Euclidean distance, i.e., shape distortion.

Fig. 4.4 shows that the upper bound of the shape distortion, derived in (4.27), contains a fixed gap with the original simulation. Therefore, we can approximate the shape distortion due to quantization with the analytical expression of (4.27).

## 4.4 Optimal Bit Allocation

The overall bit distortion takes the following form:

$$D = E\left[g^2\right] D_s + D_g \tag{4.28}$$

$$= E\left[g^2\right] C_s 2^{-\frac{2B_s}{2M-1}} + C_g 2^{-2B_g} \tag{4.29}$$

$$= \bar{C_s} 2^{-\frac{2B_s}{2M-1}} + C_g 2^{-2(B-B_s)} \tag{4.30}$$

The entries of the original channel matrix, $\mathbf{H}$, was assumed to follow a Gaussian distribution. Now, using the relationship between the singular values of a Gaussian matrix and the eigenvalues of its corresponding Wishart matrix, $E\left[g^2\right] = E[\lambda_e] = \tilde{\lambda_e}$. Here $\lambda_e$ and $\tilde{\lambda_e}$ denote the $e^{\text{th}}$ eigenvalue and the mean of the $e^{\text{th}}$ eigenvalue of the Wishart matrix respectively. $\tilde{\lambda_e}$ for different eigenvalues can be found in [54]. Since $E\left[g^2\right] = \tilde{\lambda_e}$ is also a constant with respect to bit allocation, we assumed $\bar{C_s} = C_s E\left[g^2\right]$ in (4.30). Therefore,

the optimal bit allocation problem can be defined as follows,

$$min_{B_s,B_g} \bar{C}_s 2^{-\frac{2B_s}{2M-1}} + C_g 2^{-2(B-B_s)} \tag{4.31}$$

*Lemma 3:*

The optimal bit allocation problem has the following form,

$$B_s = \frac{2M-1}{2M}B + \frac{2M-1}{4M}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right) \tag{4.32}$$

$$B_g = \frac{1}{2M}B - \frac{2M-1}{4M}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right) \tag{4.33}$$

Here, $\bar{C}_s$ and $C_g$ are the terms defined in the previous subsections.

*Proof:* See Appendix B.

Now, defining $R = B/M$ as the bit rate, i.e., bit per transmit antenna, we find,

$$R_s = \frac{2M-1}{2M}R + \frac{2M-1}{4M^2}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right) \tag{4.34}$$

$$R_g = \frac{1}{2M}R - \frac{2M-1}{4M^2}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right) \tag{4.35}$$

Here, $R_s$ and $R_g$ denotes the shape bit rate and gain bit rate respectively. Asymptotically, as the number of transmit antennas goes to infinity,

$$R_s \approx \frac{2M-1}{2M}R \tag{4.36}$$

$$R_g \approx \frac{1}{2M}R \tag{4.37}$$

The analytical expressions of (4.36) and (4.37) can be intuitively explained as follows: *The norm of a* $\mathbb{C}^M$ *vector varies across a one dimensional line. However, the shape of a* $\mathbb{C}^M$ *vector is uniformly distributed in the surface of a* $(2M-1)$ *dimensional hypersphere. Therefore, given* $2M$ *number of bits to quantize a vector, one should expend approximately* $1$ *and* $(2M-1)$ *bit to quantize the gain and shape of the vector respectively.*

Fig. 4.5 shows the effect of bit allocation in the quantization distortion of a 2x1 $\mathbb{C}^M$ MISO channel. We used (4.13) and (4.27) as the gain and shape distortion equation to find the optimal bit allocation. Here, we had 16 bits in total to quantize the vector. The

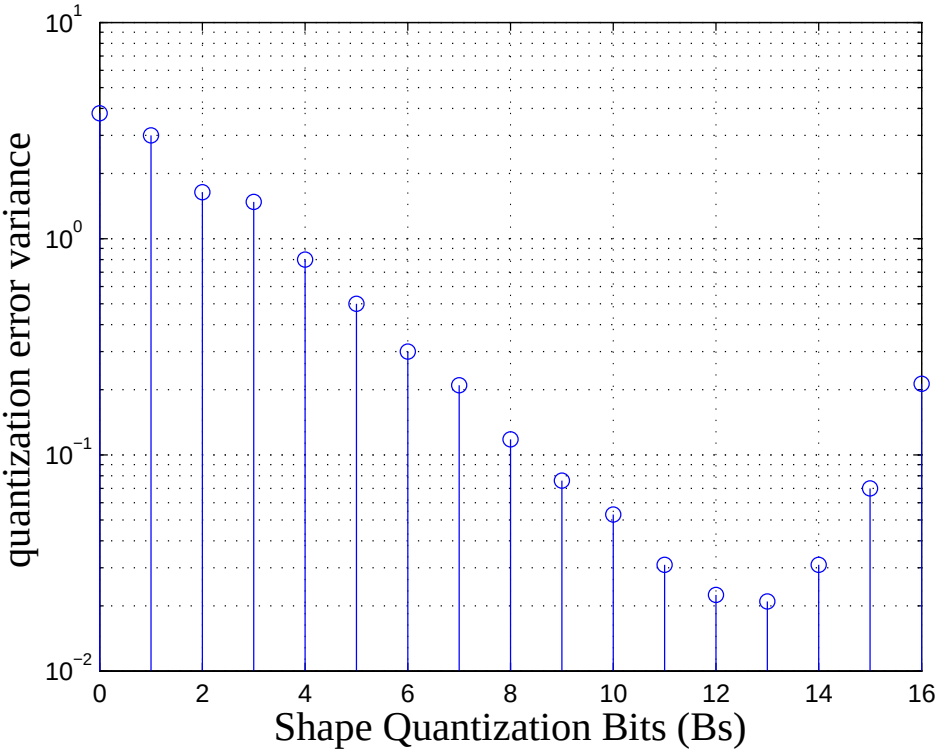



Figure 4.5: Effect of bit allocation in the 16 bit quantization of a 2x1 complex MISO channel

x axis shows the amount of bits allocated in shape quantization. Now, $B_g = B - B_s$. Therefore, the point $B_s = c$ indicates that $c$ and $16 - c$ bits were used to quantize the shape and the gain respectively where $c \in [0, 16]$. According to Fig. 4.5, the lowest distortions takes place at $B_s = 13$ ($\sigma_E^2 = 0.021$) and at $B_s = 12$ ($\sigma_E^2 = 0.023$). Our analytical expressions in (B.44) and (B.45) lead to the following optimal point: $B_s = 12.4$, $B_g = 3.6$. Thus, the predictions of our analytical expressions turn to be very close to the actual bit allocation problem.

Therefore, the quantization error for a fixed bit rate takes the following forms,

$$\begin{aligned} D =& \bar{C}_s 2^{-\frac{2B_s}{2M-1}} + C_g 2^{-2B_g} \\ =& \bar{C}_s 2^{-\frac{2}{2M-1}\left(\frac{2M-1}{2M}B + \frac{2M-1}{4M}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right)\right)} + C_g 2^{-2\left(\frac{1}{2M}B - \frac{2M-1}{4M}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right)\right)} \end{aligned} \tag{4.38}$$

$$=2^{-\frac{B}{M}} \log_2 \left( \frac{\bar{C}_s}{C_g(2M-1)} \right) \left( \bar{C}_s 2^{-\frac{1}{2M}} - C_g 2^{-\frac{2M-1}{2M}} \right) \tag{4.39}$$

$$=D_c 2^{-\frac{B}{M}} \tag{4.40}$$

Here, $D_c = \log_2 \left( \frac{\bar{C}_s}{C_g(2M-1)} \right) \left( \bar{C}_s 2^{-\frac{1}{2M}} - C_g 2^{-\frac{2M-1}{2M}} \right)$ is a constant that does not depend on the feedback overhead.

## 4.5 Overall Algorithm

We used eigenbased combining here for quantization purposes i.e., each user would project its MIMO channel into its most dominant eigenvector. Unlike Chapter 3, we do not use an MMSE receiver to observe the effect of bit allocation i.e. we use the eigenbased combiner also as the actual receiver. The steps of the overall algorithm of the proposed method are:

1. The receivers generate gain codebook of $B_g$ bits by generating dominant singular values of a random Gaussian matrix and using K-means algorithm [42]. The receivers also generate $2^{B_s}$ random unit-norm codevectors, uniformly distributed in $\mathbb{C}^M$. Both these processes are performed off-line.

2. The BS sends common pilot symbols so that receivers can estimate $\mathbf{H}_k$.

3. The receivers use $\bar{\mathbf{v}}_k = DRSV(\mathbf{H}_k)$ and find $\mathbf{F}_k = \mathbf{H}_k \bar{\mathbf{v}}_k$. Here, $DRSV[\cdot]$ means finding the Dominant Right Singular Vector of the matrix.

4. The receivers use the stored codebooks to quantize the gain and shape of $\mathbf{F}_k(:,1)$. Quantization is done based on Euclidean distance.

5. $\mathbf{Q}^{opt} = \min_{\mathbf{Q}} \left( \sigma^2 + \frac{\sigma_E^2 P_{max}}{M} \right) tr(\mathbf{J}^{-1})$, such that $tr(\mathbf{Q}) \leq P_{max}$. $\sigma_E^2 = D$ of (4.40). $\mathbf{J}$ follows (3.48).

6. $\bar{\mathbf{u}}_k = \mathbf{J}^{-1} \hat{\mathbf{f}}_k \sqrt{q_k}$, $\mathbf{u}_k = \bar{\mathbf{u}}_k / ||\bar{\mathbf{u}}_k||_2$.

7. $\mathbf{p} = \mathbf{q}$.

8. $\mathbf{v}_k = ||\bar{\mathbf{u}}_k|| \times \bar{\mathbf{v}}_k$.

Here, $\mathbf{u}_k$,$\mathbf{v}_k$,$\mathbf{p}$ and $\mathbf{q}$ denote the same things as in Chapter 3. The reasoning for using

$||\bar{\mathbf{u}}_k||$ in the receiver design is as follows: $\bar{\mathbf{u}}_k$ of step 6 minimizes the virtual uplink SMSE of the system. However, $\bar{\mathbf{u}}_k$ is normalized to limit the overall transmitted power. Therefore, $||\bar{\mathbf{u}}_k||$ is used as a post-multiplication factor at the receiver end to scale the received signal i.e. to minimize the overall SMSE [50, 51]. $||\bar{\mathbf{u}}_k||$ impacts the performance of pulse amplitude modulation based system. The BS can inform the receivers regarding their individual $||\bar{\mathbf{u}}_k||$ by expending a few bits in a limited feedforward path.

## 4.6 Numerical Simulations

We present simulations to observe the effect of bit allocation in a multiuser multiantenna system. In our multiuser system model, the base station has two transmit antennas and there are two receivers. Each receiver has 2 receive antennas and receives 1 data stream. The feedback overhead per user is 16 bits.

Fig. 4.6 shows the effect of bit allocation on the quantization error of the mentioned vector. Figure 4.6 shows that $B_s = 13$ and $B_g = 3$ provides the optimal bit allocation. The derived equations in (4.32) and (4.33) lead to the following result: at the optimal point, $B_g = 2.6$, $B_s = 13.4$. Therefore, the theoretical solution matches closely with the simulated result.

We observe the effect of bit allocation in the SMSE performance of a 16-QAM modulation based system. Fig. 4.7 shows that $B_s = 12$ and $B_s = 13$ leads to the minimum SMSE whereas, $Bs = 16$ leads to higher SMSE. Therefore, optimal bit allocation across gain and shape feedback provides better performance in terms of SMSE.

Fig. 4.8 shows that $B_s = 12$ and $B_s = 13$ performs best in terms of BER, too. Multiuser interference changes both *norm and shape* of the received signal. Since $B_s = 16$ uses $B_g = 0$, the calculation of $||\bar{\mathbf{u}}_k||$, in this case, becomes erroneous. The received signal at the antenna does not get scaled properly and the change in norm due to multiuser interference is not corrected. Therefore, the overall SMSE is not minimized. This leads

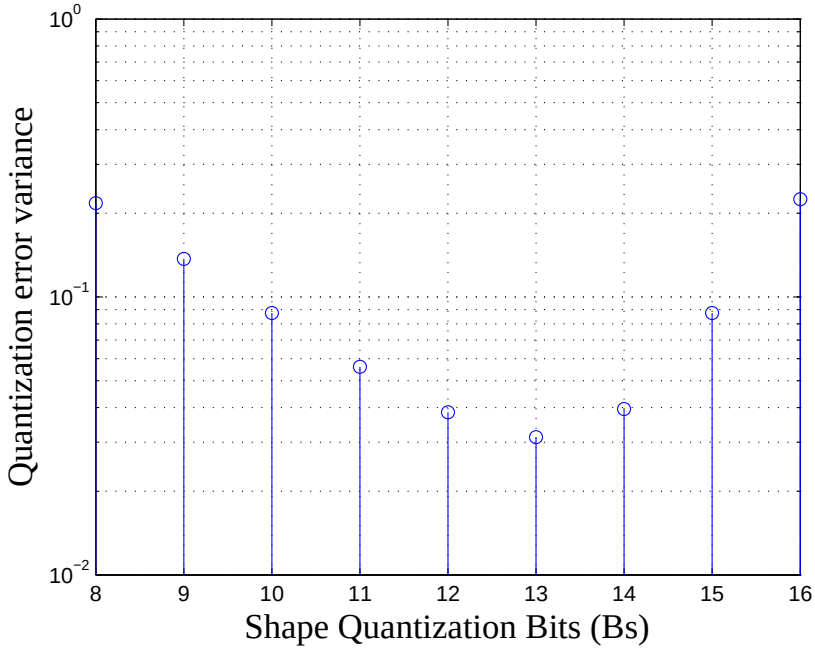



Figure 4.6: Effect of bit allocation in the quantization of the product of dominant eigenvalue & the corresponding eigenvector of a 2 x 2 MIMO channel

to the inferior performance of $B_s = 16$.

The norm of the recovered signal does not have any effect in phase shift keying (PSK) based systems. Therefore, we observe the effect of bit allocation on the BER performance of the MU-MIMO system using QPSK modulation. Here, $M = 2$, $N = [2\ 2]$, $L = [1\ 1]$, $B = 12$. The optimal bit allocation analysis leads to following result: $B_s = 10$, $B_g = 2$. However, Fig. 4.9 shows that the $B_s = 12, B_g = 0$ leads to the least BER in the QPSK modulation based system. This suggests that the use of all available bits in shape quantization leads to the best performance in QPSK system, in terms of BER. This result is in direct contrast with our optimal bit allocation derivation. We assume that this may occur due to the fact that the BER in QPSK only depends on the phase, not gain, of the recovered signal.

A closer look at the order of the performance of different bit allocation in Fig. 4.6,

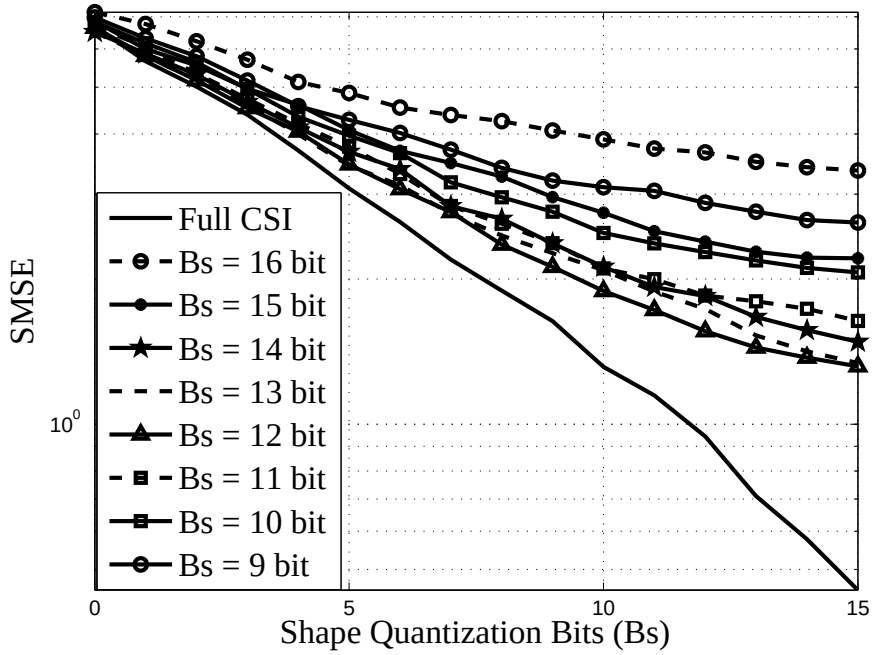



Figure 4.7: Effect of bit allocation in the SMSE of 16-QAM system, $M = 2, N = [2\ 2], L = [1\ 1], B = 16$

Fig. 4.7 and Fig. 4.8 reveal that overall quantization error is related with the SMSE & BER performance of a 16QAM modulation based system. However, since QPSK depends only on phase, the quantization error & SMSE do not reflect the BER performance of a QPSK based system. An extension of this work should be the investigation of the use of higher bit rate feedback and the inclusion of other modulation systems.

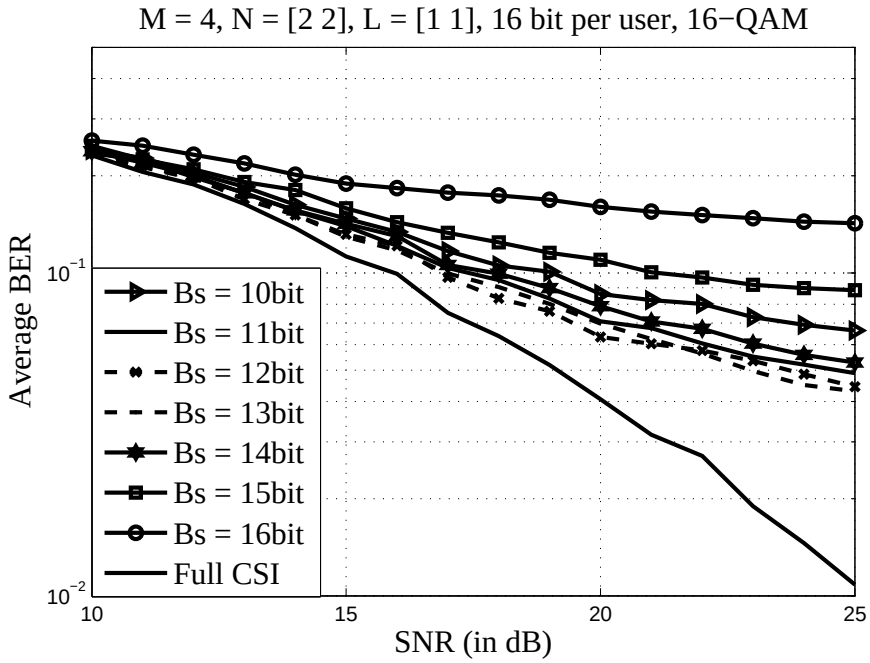



Figure 4.8: Effect of bit allocation in the BER of 16-QAM systems, $M = 2, N = [2\ 2], L = [1\ 1], B = 16$

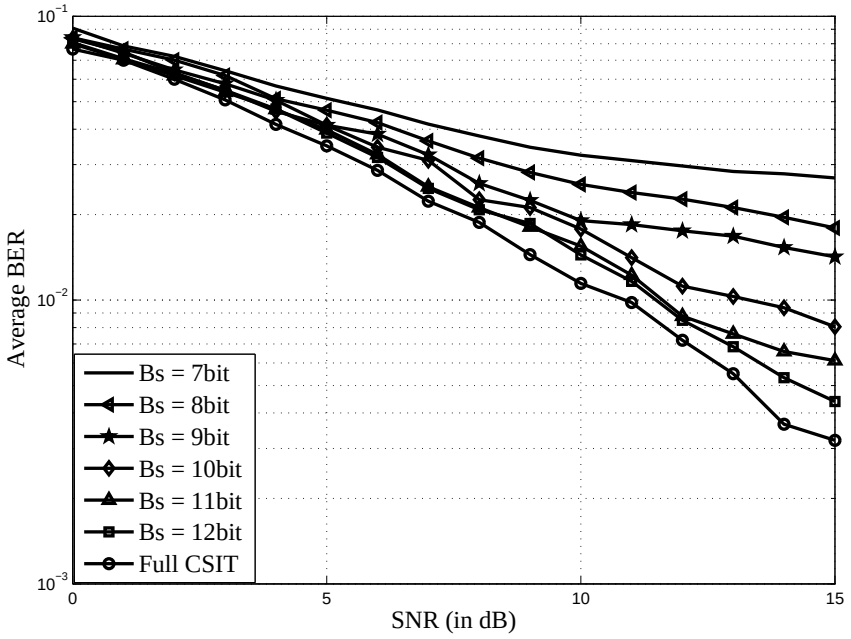



Figure 4.9: Effect of bit allocation in BER in QPSK, $M = 2, N = [2\ 2], L = [1\ 1], B = 12$

# Chapter 5

# Quantized Feedback in a Time Varying Multiuser Channel

In the previous two chapters, we assumed a block fading channel model at each channel realization. The users had to quantize the full channel information for feeding back to the base station. However, time varying channel is a more realistic model for the application under consideration and the use of past channel knowledge can lead to significant reduction in feedback overhead. This motivates us to design quantized feedback in a time varying multiuser channel. In this chapter, we design linear least squares and recursive least squares based adaptive predictors and 2 bit differential quantizers. Compared to the existing differential feedback literature, our proposed quantizer provides three advantages:

1. The controller parameters are flexible enough to adapt themselves to different vehicle speeds.

2. The model is backward adaptive i.e., the base station and receiver can agree upon the predictor and variance estimator coefficients without the explicit exchange of the parameters

3. It can outperform fixed quantizer even when the correlation between two successive

Table 5.1: Channel Parameters

| Parameter | Value(units) |
|---|---|
| Carrier Frequency ($\mathbf{f}_c$) | 2.5 GHz |
| Channel Sampling Rate ($\mathbf{f}_s$) | 200 Hz |
| Frame Duration ($\mathbf{T}_{fr}$) | 5 ms |

channel samples becomes as low as 0.05.

## 5.1 System and Feedback Model

We use the same system model and the notations that were enlisted in Chapter 2 and 3. However, since we assume time varying channels in our current scheme, our overall system model incorporates channel mobility and feedback model.

The channels are temporally correlated and assumed to follow a modified version of Jakes' model [59]. Here, channel parameters are selected (as in Table 5.1) to represent typical values of the WiMax standard [1].

In this chapter, we use two different feedback methods. We introduce those two methods here. The corresponding algorithms will be described in details in the respective sections.

**Channel Quantization**

In this method, full channel knowledge is quantized and sent back to the base station (BS). The receivers expend 2 bits on the differential quantization of the real and imaginary parts of each channel entry based on minimum Euclidean distance. The base station (BS) uses the following channel model:

$$\mathbf{H} = \hat{\mathbf{H}} + \tilde{\mathbf{H}} \tag{5.1}$$

Here, $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ denote the quantized channel and error in channel feedback respectively.

This channel model is used to find the optimal $\mathbf{F}$ through iteration between $\mathbf{V}$ and $\mathbf{Q}$ [32]. This method allows the implementation of an optimal receiver at the expense of higher feedback overhead.

**Eigenentry Quantization**

Here, we use the feedback model proposed in Chapter 3 i.e., the individual users only feed back the effective vector downlink channel knowledge. At first, let us reiterate the singular value decomposition model of the channel i.e.,

$$\mathbf{H}_k = \mathbf{A}_k \Sigma_k \mathbf{B}_k^H \tag{5.2}$$

In this proposed method, for the purposes of quantization only, the receivers use, as $\mathbf{V}_k$, the $L_k$ right singular vectors corresponding to the maximum singular values of $\mathbf{H}_k$. So, $\mathbf{V}_k = \mathbf{B}_k(:, 1 : L_k)$. Therefore, $\mathbf{F}_k = \mathbf{A}_k(:, 1 : L_k) \times \mathbf{\Sigma}_k(:, 1 : L_k)$. Thus, each receiver projects its own MIMO channel into the set of dominant eigenvectors.

The receivers expend 2 bits to perform adaptive differential quantization of each real and imaginary entry of $\mathbf{A}_k(:, 1 : L_k)$ and fixed scalar quantization of each entry of $\mathbf{\Sigma}_k(:, 1 : L_k)$. Since $\mathbf{F}$ is quantized and sent to the base station, the transmitter assumes the following channel model,

$$\mathbf{F} = \widehat{\mathbf{F}} + \widetilde{\mathbf{F}} \tag{5.3}$$

Here, $\widehat{\mathbf{F}}$ and $\widetilde{\mathbf{F}}$ represent the quantized effective channel and error in the feedback respectively. The dimension of $\mathbf{F}$ is $M \times L$. As explained in Chapter 3, since $L \leq N$, this method saves feedback overhead at the expense of a sub-optimal receiver.

## 5.2 Adaptive Differential Quantizer Model

The receiver uses the adaptive differential quantizer model of Stroh [53], shown in Fig. 5.1, originally proposed for speech processing. The left and right sides of the channel block

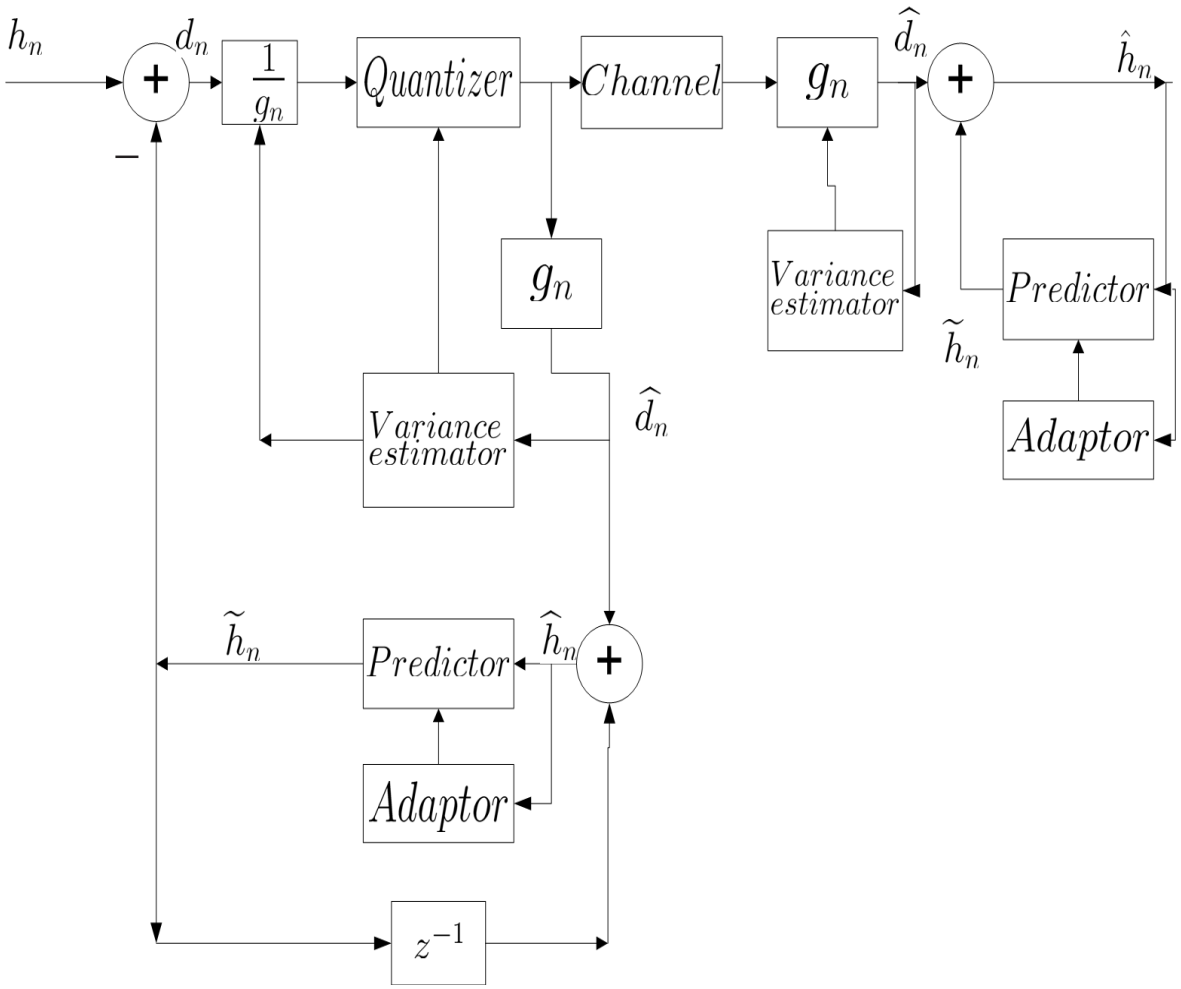



Figure 5.1: Block diagram of the adaptive differential quantizer

are located at the receiver and base station respectively. Since we assume the channel variance to be Gaussian, a unit variance 2 bit Gaussian quantizer [39, 43] is used in the quantizer block. Let $h_n$ and $\hat{h}_n$ represent the original and quantized channel parameters at the $n^{\text{th}}$ instant; $\tilde{h}_n$ denotes the predicted channel entry, calculated based on the past samples of $\hat{h}_n$.; $d_n = h_n - \tilde{h}_n$ is the difference signal between the incoming channel entry $h_n$ and predicted channel parameter $\tilde{h}_n$; $g_n$ is used to normalize the variance of the difference signal i.e., to avoid granular noise and overloading; $\hat{d}_n = d_n + qn_n$ where $qn_n$ is the quantization noise at the $n^{\text{th}}$ instant; $z^{-1}$ denotes a one sample delay. Note that we assume that the receiver knows the CSI exactly.

To focus on quantization, we assume delay-less and noise-free feedback in our work. Therefore, using the symmetry of the receiver and base station in adaptive differential coding, $\hat{h}_n = h_n + qn_n$ [16,53]. Here, the adaptor block controls the predictor coefficients and the variance estimator block estimates $g_n$. The predictor coefficients and variance estimator parameters depend on $\hat{h}_n$ and $\hat{d}_n$, rather than on $h_n$ and $d_n$. Therefore, unlike the differential feedback model proposed in [24,35,36], the BS can reproduce the predictor and variance estimator parameters without the explicit transmission of the coefficients.

## 5.3 System Design

We design the predictor and variance estimator blocks to control the performance of the adaptive differential quantizer. The three-fold objectives of the system design are given below:

1. The system needs to be backward-adaptive i.e., the control parameters depend on $\hat{h}_n$ and $\hat{d}_n$. This alleviates the need of explicit exchange of control parameters for different channel correlations.
2. Minimize the quantization error variance.
3. Minimize the transient time.

We use both linear least squares (LLS) and recursive least squares (RLS) based predictor and variance estimators in our work.

### 5.3.1 LLS Based Predictor

The LLS predictor uses the model

$$\tilde{h}_n = \sum_{j=1}^{T} w_{j,n} \hat{h}_{n-j} \tag{5.4}$$

Here, $T$ is the predictor order and $\mathbf{w}_{j,n}$ is the $j^{\text{th}}$ weight coefficient at the $n^{\text{th}}$ time instant. The predictor coefficients are computed to minimize the mean squared error

$$\epsilon^2 = \frac{1}{Lp}\sum_{i=1}^{L_p}\left[\hat{h}_{n-i} - \sum_{j=1}^{T} w_{j,n}\hat{h}_{n-i-j}\right]^2 \tag{5.5}$$

Here, $L_p$ is the learning period. $\epsilon^2$ is the fitting or average prediction error. The weights are the solution to the linear equation,

$$\mathbf{\Phi}(\mathbf{n})\mathbf{w}(\mathbf{n}) = \psi(\mathbf{n}) \tag{5.6}$$

where $\mathbf{w}(\mathbf{n})$ is the $T \times 1$ vector of predictor coefficients, at the $n^{\text{th}}$ time instant; $\mathbf{\Phi}(\mathbf{n})$ is the $T \times T$ covariance matrix estimate and $\psi(\mathbf{n})$ is a $T \times 1$ vector given by,

$$\mathbf{\Phi}(\mathbf{n})_{j,k} = \sum_{i=1}^{L_p}\hat{h}_{n-i-j}\hat{h}_{n-i-k} \tag{5.7}$$

$$\psi(\mathbf{n})_{j,1} = \sum_{i=1}^{L_p}\hat{h}_{n-i}\hat{h}_{n-i-j} \tag{5.8}$$

Clearly, this has the same form as the Weiner filter with a finite training period.

For LLS, the factor normalizing the variance, $g_n$, is given by:

$$g_n = k\sqrt{\frac{1}{L_R}\sum_{i=1}^{L_R}\hat{d}^2_{n-i}}. \tag{5.9}$$

Here, $L_R$ is the learning period of the variance estimator. $k$ is a constant which is used to compensate for the bias in the estimate. Since, $\hat{d}_n$ is corrupted by quantizer noise, $k$ must be chosen via experiment.

### 5.3.2 RLS Based Predictor

As we will show later, linear least square based predictors perform close to ideal Weiner filter predictors. However, reducing the steady state error to acceptable levels requires increasing the learning period and an attendant increase of transient time [53]. The

increase of transient time in LLS based differential quantizer motivates investigation of a recursive least square based backward predictor [22]:

$$\mathbf{\Phi}(\mathbf{n}) = \sum_{i=1}^{n-1} \lambda^{n-1-i} \hat{\mathbf{h}}(i) \hat{\mathbf{h}}^H(i) \tag{5.10}$$

$$\psi(\mathbf{n}) = \sum_{i=1}^{n-1} \lambda^{n-1-i} \hat{\mathbf{h}}(i) \hat{d}^H(i) \tag{5.11}$$

$$\mathbf{\Phi}(\mathbf{n}) = \lambda \mathbf{\Phi}(\mathbf{n-1}) + \hat{\mathbf{h}}(n-1) \hat{\mathbf{h}}^H(n-1) \tag{5.12}$$

$$\psi(\mathbf{n}) = \lambda \psi(\mathbf{n-1}) + \hat{\mathbf{h}}(n-1) \hat{d}^H(n-1) \tag{5.13}$$

$$\mathbf{\Phi}(\mathbf{n}) \mathbf{w}(\mathbf{n}) = \psi(\mathbf{n}) \tag{5.14}$$

Here, $\hat{\mathbf{h}}(i) = \left[\hat{h}_i, \cdots, \hat{h}_{i-T+1}\right]$ and $\lambda$ is the memory factor of the predictor.

For the RLS variance estimator, $g_n$ is calculated as

$$v_n = \sum_{i=1}^{n-1} k_2^{n-1-i} \hat{d}_i^2 \tag{5.15}$$

$$g_n = k_1 \sqrt{\frac{1}{\sum_{i=1}^{n-1} k_2^{n-1-i}} v_n} \tag{5.16}$$

As $n$ becomes large,

$$g_n = k_1 \sqrt{(1-k_2)\left(k_2 v_{n-1} + \hat{d}_{n-1}^2\right)}. \tag{5.17}$$

Here, $k_1$ plays the same role as $k$ in LLS variance estimator. $k_2$ is the memory factor of the variance estimator.

## 5.4 Selection of channel parameters

Fig. 5.2 shows the effect of the variance estimator learning period ($L_R$) in the quantization error variance of LLS based feedback. As Fig. 5.2 shows, the quantization error variance decreases as learning period increases. However, the increase of learning period in the variance estimator also increases the transient time.

Fig. 5.3 shows the effect of learning period of the predictor ($L_p$) in LLS based feedback. The effect of the learning period length in the predictor block follows the same pattern

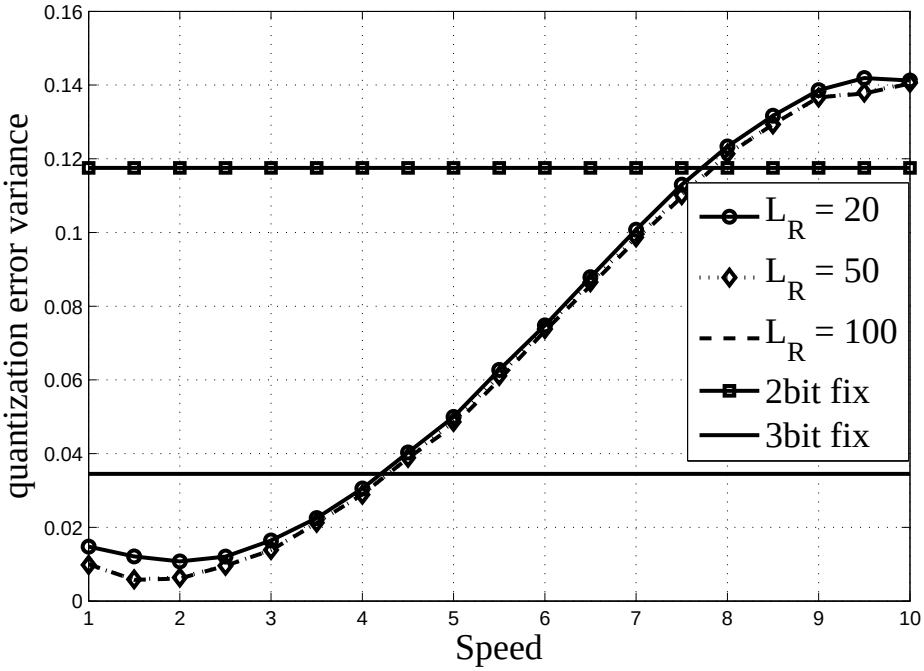



Figure 5.2: Effect of learning period of the variance estimator $L_R$ in the quantization error variance of LLS based feedback

of the learning period length of the variance estimator block. The increase in learning period comes with the decrease in quantization error and increase in transient time. We heuristically choose a learning period of 100 samples in our algorithm. Since, the time duration between two successive sample is 5 ms [1], this will lead to a transient time of around 500 ms.

Fig. 5.4 shows the effect of LLS bias ($k$) in the quantization error performance of the system. As Fig. 5.4 shows, a lower value in the bias leads to a relatively small quantization error variance at high speed. However, it leads to a relatively high quantization error variance at low speed. The higher bias values tend to show the opposite pattern. We stick to a bias value of 1.1 which provides a trade-off between these two extremes.

The predictor memory constant and memory factor, in the given range, do not have a significant impact on the quantization error variance. We use a predictor memory

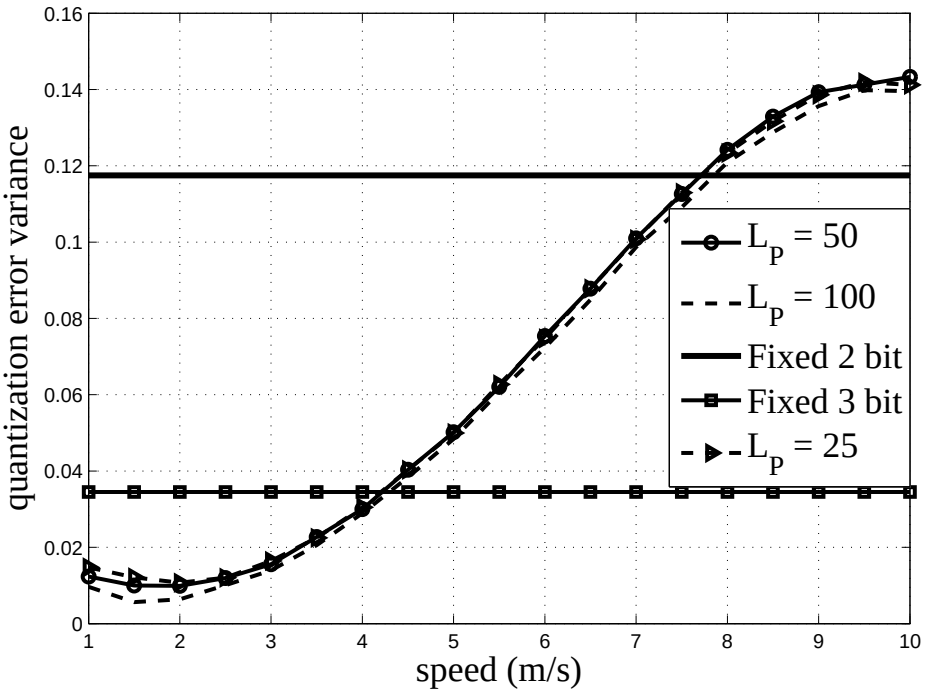



Figure 5.3: Effect of learning period of the predictor ($L_P$) in the quantization error variance of LLS based feedback

constant of 0.98 and a memory factor of 0.9 in our work. The design parameters used in our experiments are provided in Table. 5.2.

**Linear precoding algorithm with channel quantization based feedback**

Given the approach taken above, the overall precoding algorithm is as follows: The BS designs the linear precoding algorithms using the adaptive differential quantizer. Since the BS has quantized knowledge of all the entries of $\hat{\mathbf{H}}$, the linear transceiver can be designed with a few changes to the algorithm of [32]. Briefly, the resulting algorithm is:

1. The BS sends common pilot symbols so that receivers can estimate $\mathbf{H}_k$.

2. The receivers feed back each real and imaginary entry of $\hat{\mathbf{H}}_k$ using the proposed adaptive differential quantizer.

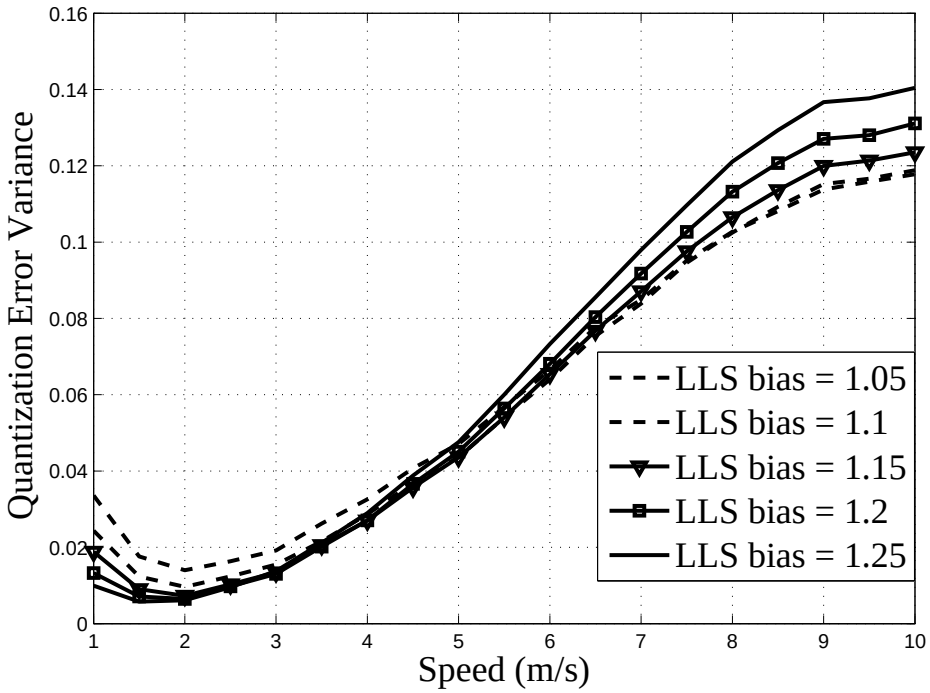



Figure 5.4: Effect of variance estimator bias ($k$) in the quantization error variance of LLS based feedback

3. The BS initializes $\mathbf{V}_k = SVD(\hat{\mathbf{H}}_k)$ and $q_k = 1, \forall k$.

4. Let, $\mathbf{J} = \hat{\mathbf{H}}\mathbf{V}\mathbf{Q}\mathbf{V}^H\hat{\mathbf{H}}^H + \sigma^2\mathbf{I}_M + \sigma^2_{E_H} tr[\mathbf{Q}]\mathbf{I}_M$. The BS iterates between optimum $\mathbf{V}$ and $\hat{\mathbf{Q}}$ to minimize $\mathbf{J}$ [32].

5. $\mathbf{U}_k = \mathbf{J}^{-1}\mathbf{H}_k\mathbf{V}_k\sqrt{Q}_k$.

6. $\mathbf{p} = \mathbf{q}$.

7. The BS sends dedicated pilot symbols and the users implement MMSE receivers:

$$\mathbf{V}_k = \left(\mathbf{H}_k^H\mathbf{U}\mathbf{P}\mathbf{U}^H\mathbf{H}_k + \sigma^2\mathbf{I}\right)^{-1}\mathbf{H}_k^H\mathbf{U}_k\sqrt{\mathbf{P}}_k. \tag{5.18}$$

$\sigma^2_{E_H}$ is the quantization error variance associated with each scalar entry of the matrix $\tilde{\mathbf{H}}\mathbf{V}$. We assumed $\mathbf{V}$ to be deterministic here to simplify the quantization error analysis.

Table 5.2: Design Parameters

| Parameter Description | Parameter Notation | Value |
|---|---|---|
| Learning Period of the predictor in LLS | $L_P$ | 100 |
| Learning Period of the variance estimator in LLS | $L_R$ | 100 |
| Bias of the variance estimator in LLS | $k$ | 1.15 |
| Order of the predictor memory | $T$ | 2 |
| Memory factor of the predictor in RLS | $\lambda$ | 0.98 |
| Memory factor of the variance estimator in RLS | $k_2$ | 0.9 |
| Bias of the variance estimator in RLS | $k_1$ | 1.1 |

We used $g_n^2$ as the quantization error variance in section 5.2 i.e., our proposed differential channel model. Since we perform quantization of the real and imaginary component independently, $\sigma_{E_H}^2$ can be readily computed as $g_{nr}^2 + g_{ni}^2$ where $g_{nr}^2$ and $g_{ni}^2$ denotes the quantization variance associated with the real and imaginary part of the channel respectively.

## 5.5 Comparison of RLS and LLS based feedback

To simulate the performance of the adaptive differential quantizer, 2000 time correlated zero-mean unit-variance complex Gaussian scalar channels were generated using the channel model of [59] for each speed. In this section, we provide the comparison of RLS and LLS based feedback. Here, we choose the combination of the design parameters that, to the best of our knowledge, provide the best performance in terms of quantization error variance and transient time reduction. This set of design parameters is given in Table 5.2. One of the prime objectives of our overall study is to reduce the feedback overhead in a time varying channel. Therefore, we show the quantization error variance of fixed

2-bit and 3-bit feedback per channel entry in our work to illustrate the advantage of the LLS and RLS based feedback. The quantization error variance of 2-bit and 3-bit fixed feedback were obtained from [43].

We also show the performance of the 2 bit ideal Gaussian differential filter in time-varying scenario. The ideal Gaussian differential feedback system should consist of the following two things:

1. The minimum mean square error variance of the ideal predictor can be found as follows [22]:

$$\sigma_{d_n}^2 = \sigma_{h_n}^2 - \psi^H \mathbf{\Phi}^{-1} \psi \tag{5.19}$$

The ideal values of $\psi$ and $\mathbf{\Phi}$ can be obtained for any given speed using Doppler fading [17].

2. The quantization error variance of the 2-bit ideal Gaussian quantizer is governed by [39]:

$$\sigma_{qn_n}^2 = 0.1175 \sigma_{d_n}^2 \tag{5.20}$$

Using (5.19) and (5.20), we find the quantization error variance of an ideal Gaussian differential feedback system. Fig. 5.5 compares the quantization error variance produced by different feedback systems. Apart from pedestrian velocities (i.e., $1 - 1.5$ m/s), the RLS and LLS based feedback system provides almost same performance in terms of quantization error reduction. Fig. 5.5 shows that the performance curves of both these feedback system cross those of 3 bit fixed feedback and 2 bit fixed feedback at 4.5 m/s ($\approx$ 16 km/hr) and at 9 m/s ($\approx$ 32 km/hr) respectively. Thus, the proposed differential quantizers reduce feedback overhead by 1 bit per channel entry up to 16 km/hr and reduce quantization error with same feedback overhead up to 32 km/hr.

Note that both differential quantizers' performance becomes inferior with respect to fixed feedback as the vehicle velocity exceeds 32 km/h. Using the parameters from Table 5.1, this speed corresponds to a maximum normalized correlation of 0.0255 between two successive channel samples. Therefore, our proposed adaptive differential feedback provides better results compared to fixed quantization as long as the temporal correlation

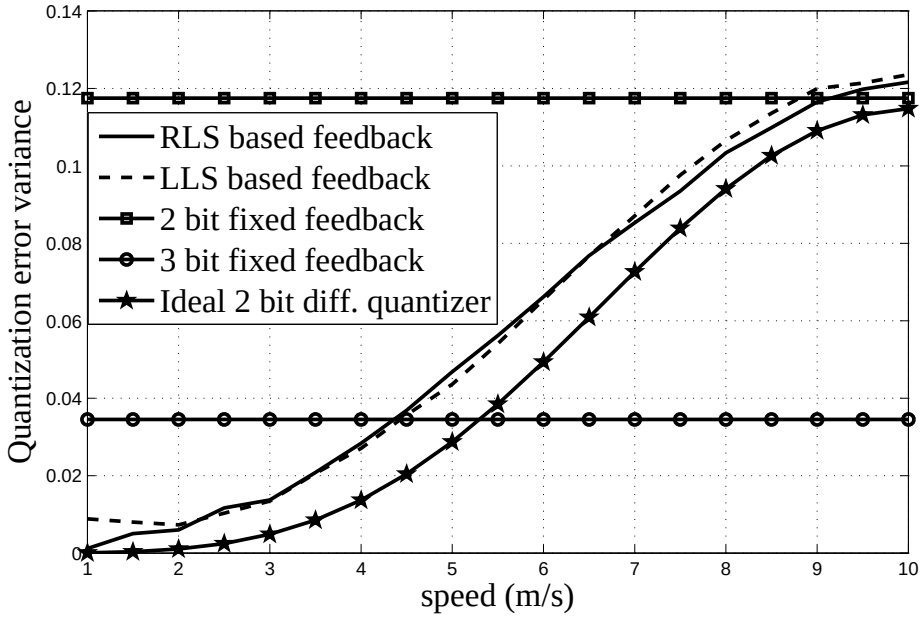



Figure 5.5: Comparison of of differential feedback with fixed feedback

between two succesive channel sample remains positive.

Fig. 5.5 shows that the ideal differential quantizer keeps almost 1 m/s ($\approx$ 3.6 km/hr) performance gap with the proposed differential quantizer. The ideal differential quantizer contains ideal co-efficients in the predictor and variance estimator block. While, our proposed feedback system updates these parameters online. If a vehicle changes its speed, our proposed feedback system can track the change of speed without explicit transmission of the coefficients. Thus, our proposed feedback model works as a more realistic feedback system at the cost of a performance gap.

Fig. 5.6 shows the transient time of RLS and LLS based feedback. The simulation was performed at 6 m/s (i.e., 21.6 km/hr). The average quantization error was calculated at every iteration. As expected, the RLS based feedback model outperforms the LLS based feedback in terms of transient time. Defining transient time to be the time when average quantization error gets reduced to 10% of its original value, the transient time of the RLS based feedback is $\approx$ 50 iterations i.e., 250 ms (since channel sampling frequency =

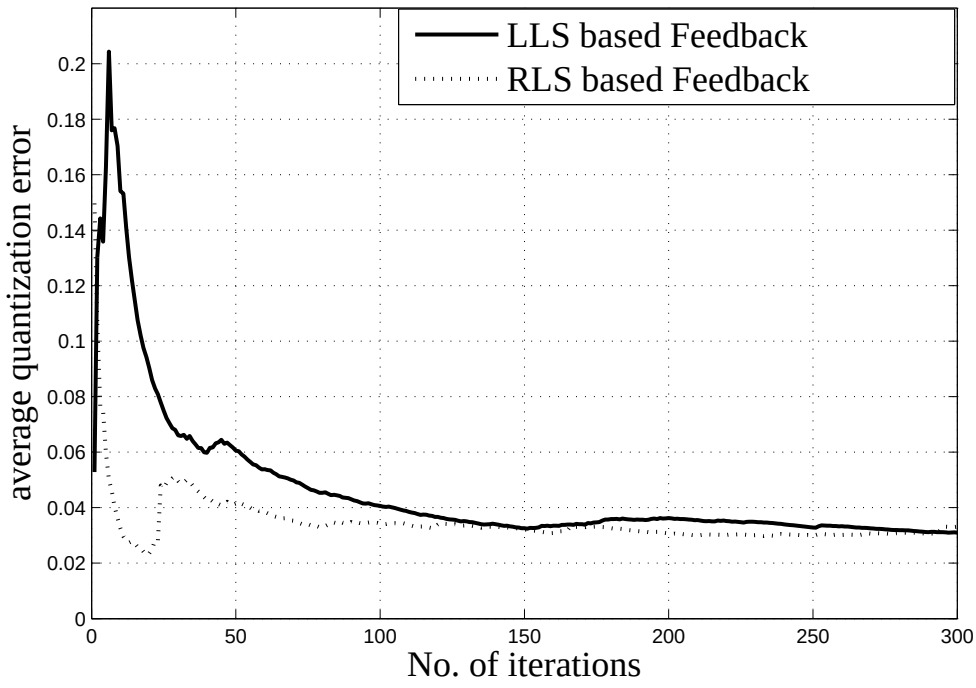



Figure 5.6: Comparison of the transient time of RLS and LLS based feedback at 21.6 km/hr

5 ms). Most of the previous works on differential feedback [24, 35, 36] assume stationary channels and need explicit exchange of control parameters. The adaptability to non-stationary channels is a major advantage of the proposed feedback model.

## 5.6 Quantization of Eigenvectors

It is now well established in the single user, single data stream, case that projecting the MIMO channel to its most dominant eigenvector yields better performance than full channel quantization with same feedback overhead [40]. Due to multiuser interference, this statement does not readily hold in multiuser multiple data stream case. However, we still investigate the performance of the adaptive differential eigen-vector feedback in multiuser time varying channels.

Table 5.3: Codebook of scalar entries of eigen-matrix

| $M = 2$ | $M = 3$ | $M = 4$ | $M = 8$ | Standard Gaussian |
|---|---|---|---|---|
| -1.34 | -1.40 | -1.43 | -1.48 | -1.51 |
| -0.43 | -0.44 | -0.45 | -0.45 | 0.45 |
| 0.43 | 0.44 | 0.45 | 0.45 | 0.45 |
| 1.34 | 1.40 | 1.43 | 1.48 | 1.51 |

The scalar entries of $\mathbf{A}_k(:, 1 : L_k)$, the left singular vector matrix of $\mathbf{H}_k$, can be adaptively differentially quantized using the same model shown in Fig. 5.1. Note that, both the adaptive predictors proposed in the previous section do not assume any particular model of the signal; they try to find the "best" predicted value based on the past observations. Therefore, if we can show the entries of $\mathbf{A}_k$ to be approximately Gaussian, the model of Fig. 5.1 can be readily applied to track $\mathbf{A}_k$. We use the following properties.

*Property 1:* The matrix of singular vectors of a rectangular Gaussian matrix is called a Haar matrix. If $\mathbf{A}$ is a $\mathbb{C}^{M\times M}$ Haar matrix, then $E\left[|\mathbf{A}_{ij}|^2\right] = \frac{1}{M}\, 1 \leq i, j \leq M$ [23].

*Property 2*: The probability distribution of $\sqrt{M}$ times the Haar matrix $\mathbf{A}$, approaches the standard complex Gaussian measure as $M \to \infty$. (4.2.11 of [45])

In practice, the entries of the Haar matrix approach a Gaussian random variable for small values of $M$. To show this, we set $N = 2$ and choose different numbers of transmit antennas, $M$. We generate $10^5$ random Gaussian distributed channels, $\mathbf{H} \in \mathbb{C}^{M\times N}$, and find the left singular matrix of $\mathbf{A} \in \mathbb{C}^{M\times M}$. We randomly pick different entries of $\mathbf{A}$. After normalizing the samples using Property 1, we find the 2 bit codebook of the collected samples using Kmeans clustering [38]. In Table 5.3 we compare the codebook with that of a unit variance 2-bit standard Gaussian quantizer [39].

In Table 5.3, the columns list the 4 level codebook, based on the scalar entries of the eigenmatrix with $M$ transmit antennas. The table shows that even for small number of

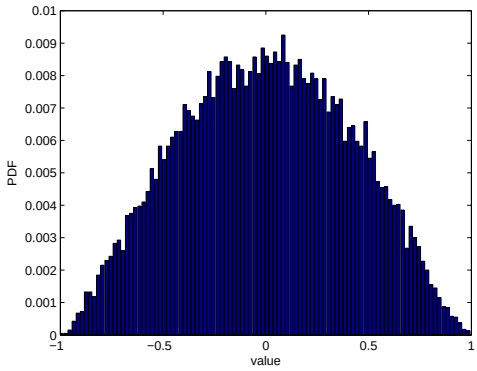



Figure 5.7: Histogram of the left singular matrix entries of a 3x2 channel

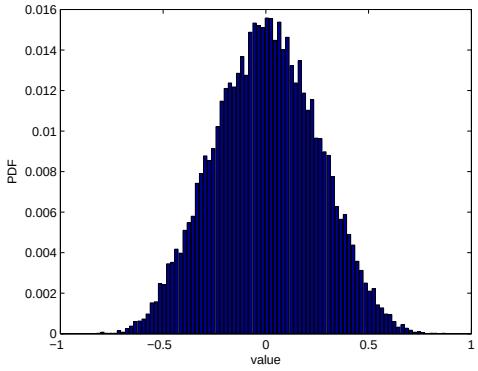



Figure 5.8: Histogram of the left singular matrix entries of a 8x2 channel

transmit antennas (e.g., 3), the normalized probability distribution of the scalar entries of the Haar matrix resemble the Gaussian distribution. Fig. 5.7 and 5.8 show that the entries of the left singular matrix of the MIMO channel appear increasingly Gaussian as the number of transmit antenna increases. This prompts us to use our proposed Gaussian differential filter to feed back the eigen entries to the base station. We spend 2 bits to quantize each real and imaginary entry of the singular vector. Since the adaptive differential quantization might change the norm of the eigen vector, the BS normalizes the eigen vector after receiving the feedback entries.

Note that the degrees of freedom of the Haar unitary matrix i.e., matrix of singular

vectors, is less than the total number of real and imaginary entries. The minimum number of parameters to represent the Haar matrix can be extracted through Givens' rotations [47]. In the literature, an adaptive controller for step size changes of Givens' rotated parameters has only been provided for pedestrian velocities [47]. The phases and Givens' rotated angles are not Gaussian distributed and least square based predictors are not optimum to track these parameters. Therefore, we stick to our proposed adaptive differential quantization policy. This ensures greater flexibility of our model to adapt to different vehicle speeds at the cost of higher feedback overhead.

### 5.6.1 Fixed quantization of singular value

In Chapter 4, we provided the eigenvalue distributions of a Wishart matrix. For moderate sizes of a MIMO system, the eigenvalues are not Gaussian distributed. Therefore, the proposed differential quantizers do not perform well in the adaptive tracking of the eigenvalues. This leads us to perform fixed quantization of the eigenvalues of the matrix using the standard Lloyd-Max quantizer [39]. The distribution of the eigenvalues is provided in Appendix B.

In this approach, receivers feed back $\hat{\mathbf{F}} = \hat{\mathbf{H}}\hat{\mathbf{V}}$ to the BS, instead of providing $\hat{\mathbf{H}}$. This saves feedback overhead as long as $L \leq N$. We used the transceiver design algorithm of Chapter 3 for this feedback. The algorithm of Chapter 3 is again summarized here:

1. BS sends common pilots to the users so that each user can estimate its own channel.

2. Each user feeds back the entries of the dominant singular vectors using the proposed adaptive differential quantizer and dominant singular values using fixed quantization.

3. Virtual uplink power allocation: Let $\mathbf{J} = \hat{\mathbf{F}}\mathbf{Q}\hat{\mathbf{F}}^H + \sigma^2\mathbf{I}_M + \sigma_{E_F}^2 tr(\mathbf{Q})\mathbf{I}_M$ and $\sigma_{E_F}^2$ be the quantization error variance of each scalar entry of $\mathbf{F}$. Then $\mathbf{Q}^{opt} = \min_{\mathbf{Q}} \left(\sigma^2 + \frac{\sigma_E^2 \sum_{i=1}^{L} q_i}{M}\right) tr(\mathbf{J}^{-1})$ s.t. $tr[\mathbf{Q}] = P_{max}; q_k \geq 0$ is convex in $\mathbf{Q}$.

4. Uplink beamforming: $\mathbf{u}_i = \mathbf{J}^{-1}\hat{\mathbf{f}}_i\sqrt{q_i}$, $||\mathbf{u}_i|| = 1$

5. Downlink power allocation: $\mathbf{p} = \mathbf{q}$.

6. BS sends dedicated pilot symbols for each of the data stream. Each user finds $\mathbf{V}_k$ using (5.18) and training.

Here, $\mathbf{F} = \mathbf{A} \times \mathbf{\Sigma}$. The BS models each scalar entry of $\hat{\mathbf{F}}$ as $\hat{\mathbf{F}}_{ij} = \hat{\mathbf{A}}_{ij}\hat{\mathbf{\Sigma}}_i$ where $\hat{\mathbf{\Sigma}}_i$ is the corresponding quantized eigenvalue of the eigenvector associated with $\mathbf{A}_{ij}$. Since the eigenvalues and eigenvector entries are quantized separately, using $\mathbf{A} = \hat{\mathbf{A}} + \tilde{\mathbf{A}}$ and $\mathbf{\Sigma} = \hat{\mathbf{\Sigma}} + \tilde{\mathbf{\Sigma}}$ in (5.3), it can be easily verified that

$$E\left[|\tilde{\mathbf{F}}_{i,j}|^2\right] = E\left[|\tilde{\mathbf{A}}_{i,j}|^2\right]\hat{\mathbf{\Sigma}}_i^2 + E\left[\tilde{\mathbf{\Sigma}}_i^2\right]\hat{\mathbf{A}}_{i,j}^2 + E\left[|\tilde{\mathbf{A}}_{i,j}|^2\right]E\left[\tilde{\mathbf{\Sigma}}_i^2\right] \tag{5.21}$$

$$E\left[|\tilde{\mathbf{F}}_{i,j}|^2\right] \approx E\left[|\tilde{\mathbf{A}}_{i,j}|^2\right]\hat{\mathbf{\Sigma}}_i^2 \tag{5.22}$$

Here, $\tilde{\mathbf{A}}_{i,j}$ and $\tilde{\mathbf{\Sigma}}_i$ are the quantization errors associated with the eigenvector and eigenvalue entry. (5.22) follows since the 1st term is much bigger than the other two terms in (5.21). The BS can estimate $E\left[|\tilde{\mathbf{A}}_{i,j}|^2\right]$ online using $g_{nr}^2 + g_{ni}^2$. Here, $g_{nr}^2$ and $g_{ni}^2$ denote the variance of the real and imaginary part of the scalar entries of the eigenvector respectively. Since the BS also knows $\hat{\mathbf{\Sigma}}_i^2$ instantly, $\sigma_{E_F}^2$ can be calculated in real time.

### 5.6.2 Discussion

Sections 5.3 and 5.6 show that, since we assumed the channel to be Gaussian and the difference of two correlated Gaussian random variables leads to another Gaussian random variable, the model shown in Fig. 5.1 provides great flexibility and can hold for different vehicle speeds. To the best of our knowledge, the proposed algorithms are the only works in the adaptive differential limited feedback literature, which can provide both the following advantages:

1. Unlike the Gauss-Markov models of [24, 35, 36], our model works when the normalized autocorrelation between successive channel samples drops below 0.5.

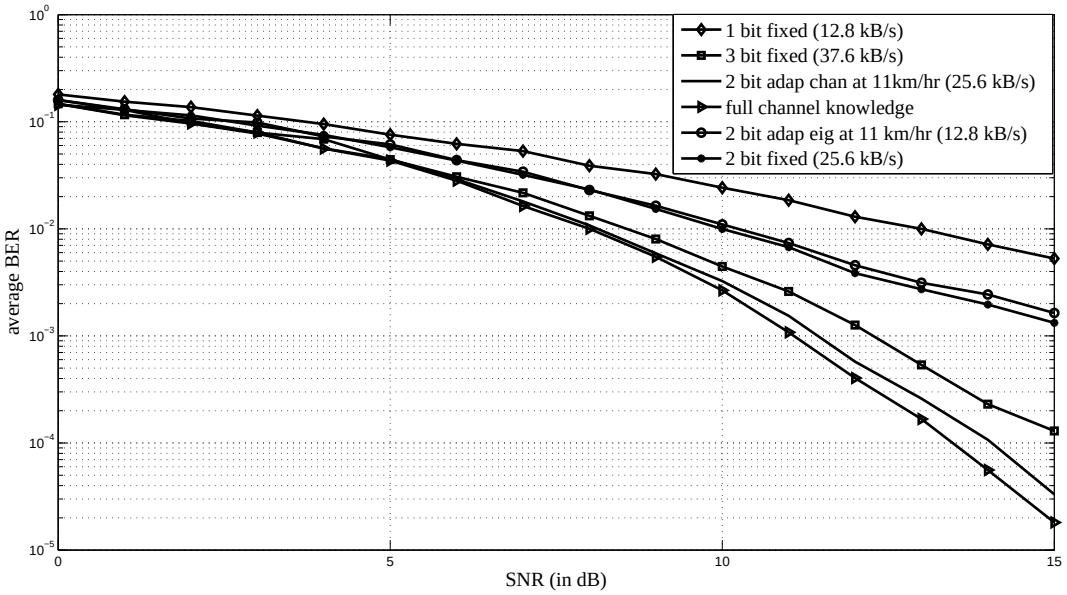



Figure 5.9: Comparing feedback methods at 11 kmhr, $M = 4, N = [4\ 4], L = [2\ 2]$

2. Unlike the feedback model proposed by [47, 56], the controlling parameters of the predictor and variance estimator in our model do not depend on the knowledge of the correlation between two successive channel samples.

## 5.7 Numerical Results

Fig. 5.9 and fig. 5.10 show the average bit error rate (BER) performances of different feedback models with their respective overheads per second at 11 km/hr speed and 32 km/hr speed respectively. We used the linear transceiver of [32] and [25] to simulate the performance of channel and eigen-matrix quantization respectively. Figure 5.9 shows that the 2-bit adaptive differential feedback outperforms 3-bit fixed feedback and performs very close to the full feedback scenario at 11 km/h . As Fig. 5.10 indicates, even at a high speed of 32 km/h (corresponds to a normalized correlation of 0.1 with a 0 degree arrival angle [17]), the proposed adaptive differential feedback reduces the BER by a factor of 2,

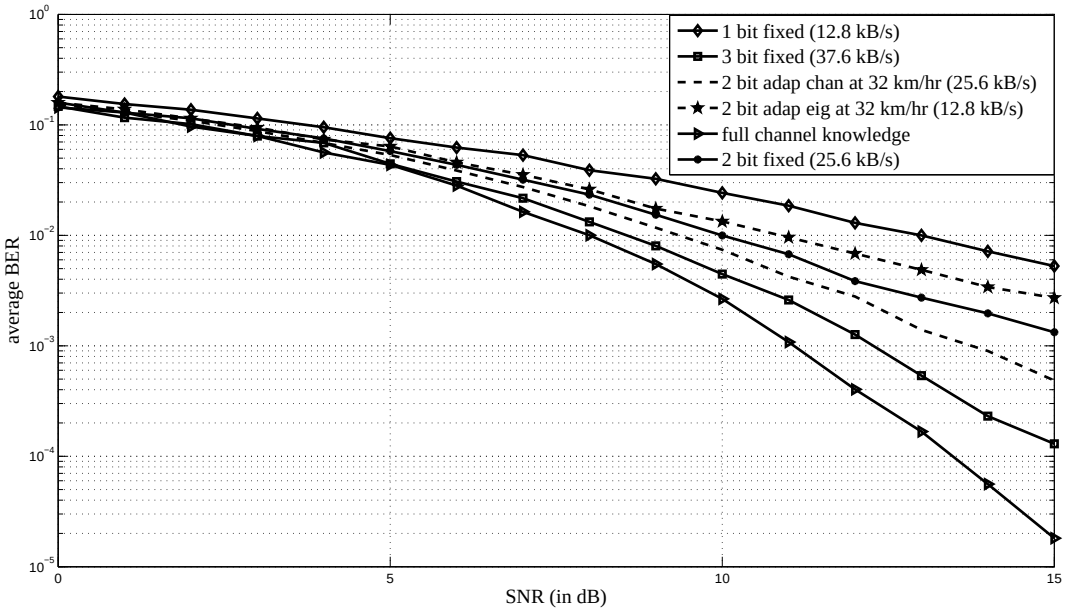



Figure 5.10: Comparing feedback methods at 32kmhr, $M = 4, N = [4\ 4], L = [2\ 2]$

with respect to 2-bit fixed feedback per channel entry i.e., with same feedback overhead.

In both Fig. 5.9 and 5.10, "2 bit adap eig" indicates the use of 2 bits to quantize each of the real and imaginary parts of the scalar entries of the eigenvector in an adaptive differential manner. Since, we assumed $N = [4\ 4]$ and $L = [2\ 2]$ in our simulation, spending 2 bits per scalar eigen-entry is equivalent to spending 1 bit per real and imaginary component of the scalar channel entry. At low speeds like 11 km/hr, the eigen-entry quantizer performs approximately as well as the 2-bit fixed quantizer and reduces the feedback overhead by a factor of 2 for almost same BER. Thus both the adaptive differential feedback methods save 1 bit per real and imaginary entry of the channel matrix at low speed (20 km/h for the channel trakcer and 8-9 km/h for the eigen tracker). This leads to a saving of $2MNF_s$ bits in feedback overhead per second. Using the channel parameters of Table 5.1 the proposed systems provide a feedback reduction of 12.8 kBit/sec.

# Chapter 6

# Conclusions and Future Work

## 6.1 Contributions

This thesis primarily focuses on different quantization algorithms of linear precoded multiuser MIMO channels that employ limited feedback. The algorithms developed in the thesis can be readily applied in frequency division duplexing systems. Since the channel reprocity between uplink and downlink does not hold in broadband time division duplexing systems, the mentioned schemes can be utilized to broadband time division duplex systems, too [19].

Chapter 3 separated the design of the transmitter and the receiver between the base station and individual user end. There have been lot of works in the literature that focus on joint transmitter-receiver optimal design at the base station. The separate design of precoding and decoding matrix allowed low feedback overhead at the expense of a sub-optimal receiver. The proposed subobtimal algorithm was shown to outperform the optimal methods with quantized channel knowledge. The major novel contribution of Chapter 3 lies in the extension of the maximum expected signal combining (MESC) to the multiple data streams per user scenario. This algorithm retains the benefits of eigenbased combining and quantization based combining at low and high SNR respectively. Our

proposed algorithm was shown to outperform several other available linear precoding based MU MIMO systems.

Chapter 4, at first, derived the quantization error variance of the gain and shape of a vector for a given number of feedback bits. Thereafter, it provided optimal bit allocation across different eigenvalues and eigenvectors of a matrix. This led to the reduction of the overall SMSE of a multiuser MIMO system based on limited feedback. The derived algorithms can be readily applied to MU MIMO systems that use eigenbased combining. Using the norm of the effective vector downlink channel in quantization based combining (QBC), the obtained result can be applied in QBC based MU MIMO systems, too.

Chapter 5 focused on adaptive differential feedback algorithms in time varying channels. Two adaptive differential scalar quantization models were proposed in the work. They are: i) channel quantization and ii) eigenentry quantization. The differential quantzers were shown to outperform fixed quantizers as long as the correlation between two successive channel samples remained postivie. Unlike most of the previous works on differential limited feedback literature, the proposed algorithms did not require the explicit exchange of channel correlation or control parameter information between the base station and the user end. Both the algorithms were shown to provide several kBit/sec feedback overhead up to $15 - 16$ km/hr in present wireless communication standards.

## 6.2 Future Work

This thesis incorporates several limited feedback based quantization techniques in multiuser MIMO channel. We suggest that this thesis may form a basis for future research that builds upon some of the ideas expressed here. The possible areas of future research are furnished below:

1. Allocation of optimal bits across gain and shape of the channel led to superior result, in terms of BER, in 16QAM modulation based system. However, allocating the

total number of bits in the shape outperformed optimal bit allocation in terms of BER in QPSK modulation based system. More investigation is needed to explore the correlation between quantization error variance, SMSE and BER of different modulations.

2. As explained in Chapter 5, Givens' rotation allows to extract necessary and sufficient parameters to construct the Haar matrix. Therefore, the feedback overhead of scalar adaptive differential quantization can be reduced by tracking Givens rotated parameters. Due to the non-linear nature of the Givens rotated parameters, our LLS and RLS based differential quantizers could not track these. The use of non-linear filters like particle filters in the tracking of givens rotated parameters should be investigated in future.

3. Chapter 3 shows that the knowledge of quantization error variance is required in designing the precoding beamformer and power allocator matrix. The quantization error variance of EBC and QBC have already been derived in the literature. MESC converges to EBC and QBC at low and high SNR respectively. Therefore, the quantization error variance of MESC should follow that of EBC and QBC at low and high SNR respectively. The error variance of MESC at intermediate areas of SNR remains an open area of future work.

4. We only focused on multiuser MIMO systems with flat fading channels in our work. An important adjunct to this work would be the extension of this work to systems employing orthogonal frequency division multiplexing.

# Appendix A

# Quantization Error Analysis and Code book Generation

## A.1 Quantization Error Analysis

Due to the structure of the receive combining, the quantization error in the quantized feedback effective MISO channel varies from low to high SNR. Thus, the variance of $\tilde{\mathbf{f}}_i$ varies, too. In the following, we give a brief explanation of the quantization error variance in the high and low SNR scenario.

### A.1.1 Quantization Error at Low SNR

In the low SNR region, we can assume, $\sigma^2 \gg \left( \sum_{n \neq j} \frac{P}{L} \left| \mathbf{f}_{k_j}^H \mathbf{u}_{k_n} \right|^2 + \sum_{m \neq k, l \in [1, L_m]} \frac{P}{L} \left| \mathbf{f}_{k_j}^H \mathbf{u}_{m_l} \right|^2 \right)$ in (3.19). Therefore, the proposed scheme leads to maximizing signal power and the quantization problem can be formulated as finding the decoding vector that would maximize the signal power and then finding the quantized code vector that is closest to the newly formed vector downlink MISO channel.

Due to the formulation of the MSIP approach, the error variance of quantization

error, $\sigma_E^2$, is measured in terms of the angle spread between the original and quantized vectors. In [48], the quantization error of $\tilde{\mathbf{f}}$ was given in the following form,

$$\sigma_E^2 = E\left[\sin^2\left(\angle\left(\mathbf{f}_{k_j}, \widehat{\mathbf{f}}_{k_j}\right)\right)\right] \leq 2^{\frac{-B}{M-1}} \tag{A.1}$$

### A.1.2 Quantization error at high SNR

In Section A.1.1, we showed that our proposed algorithm is equivalent to QBC at high SNR for one data stream per user. Simulation results in Section 3.8 showed the simulation of the convergence of this algorithm to QBC for multiple data streams per user. Therefore, we analyze the high-SNR quantization error of our receiving combining scheme using the concepts of QBC.

When each user receives one data stream, QBC chooses the codevector with the least quantization error and thus converts a MIMO channel into an effective MISO channel [30]. The quantization error in this case is upper bounded by $2^{\frac{-B}{M-N_k}}$ [30]. Using the same notion, for a multiple data stream per user scenario, the effective MISO channel of the $j^{\text{th}}$ stream of a particular user can be chosen to generate the $j^{\text{th}}$ least quantization error with respect to its original MIMO channel. The expected quantization error of the $j^{\text{th}}$ data stream (in terms of error tolerance) of the $k^{\text{th}}$ user in this method satisfies [18, 30],

$$\sigma_E^2 \leq j \times 2^{\frac{-B}{M-N_k}} \tag{A.2}$$

Here, $j \in [1, L_k]$. Therefore, quantization error of any stream of the $k^{\text{th}}$ satisfies, $\sigma_E^2 \leq L_k \times 2^{\frac{-B}{M-N_k}}$. Note that the quantization method described in the previous passage can lead to intra-user interference due to the correlation of two codevectors of a particular codebook. Our proposed algorithm avoids this scenario by incorporating the intra-user interference in receiver combining. However, the codevectors chosen for two different streams of a user vary with time and become mutually statistically uncorrelated in the long term of multiple channel realizations. Therefore, we hypothesize that the quantization error of our algorithm matches with that given by (A.2) at high SNR.

The proposed receive combining scheme incorporates both an increase in signal power and reduction in (intra and inter user) interference. The trade-off between these two depends on the SNR. Due to the adaptive nature of this method, the expected quantization errors for intermediate SNR cases are very hard to derive. In our simulations we assumed the quantization error to take the form of (A.1) at low SNR (0 dB) and changed this value linearly with transmitted power so that it converged to the form of (A.2) at high SNR (30 dB). Investigating the expected quantization error at the intermediate SNR remains an open research problem.

In summary, the quantization error of the proposed algorithm ranges between $2^{\frac{-B}{M-1}}$ and $L_k \times 2^{\frac{-B}{M-N_k}}$.

## A.2 Mean Square Inner Product based Vector Quantization

The concept of MSIP VQ appeared in [48]. We include a brief description of MSIP VQ here to make the thesis self-sufficient.

Let $B$ be the feedback rate. The total number of codevectors, $N = 2^B$. Let $\mathbf{f}$ represent a large set of random unit norm vectors in $\mathbb{C}^M$. Design a quantizer $C$ to maximize the MSIP,

$$(\mathbf{c}_1, \cdots, \mathbf{c}_N) = \max_{C(\cdot)} E\left| < \mathbf{f}, C\left(\mathbf{f}\right) > \right|^2 \tag{A.3}$$

Here $C\left(\mathbf{f}\right) = \hat{\mathbf{f}}$ is the quantized channel.

### Nearest Neighbour Criterion

For given code vectors $(\mathbf{c}_i; i = 1, \cdots, N)$, the optimum partition cells satisfy,

$$\mathbf{R}_i = \left(\mathbf{f} \in \mathbb{C}^M : | < \mathbf{f}, \mathbf{c}_i > | \geq | < \mathbf{f}, \mathbf{c}_j > |, j \neq i\right) \tag{A.4}$$

For $i = 1, \cdots, N$; where $R_i$ is the partition cell (Voronoi region) for the $i^{\text{th}}$ codevector $\mathbf{c}_i$.

**Centroid Condition**

For a given partition $(R_i; i = 1, \cdots, N)$ , the optimum code vector is given by,
$\mathbf{c}_i$ = (principal eigenvector) of $E\left[\mathbf{f}\mathbf{f}^H | \mathbf{f} \in R_i\right]$.

The above two conditions are iterated until the MSIP $E\left| < \mathbf{f}, C\left(\mathbf{f}\right) > \right|^2$ converges.

# Appendix B

# Quantization Distortion and Bit Allocation Proofs

## B.1 Finding $||f_g(r)||_{\frac{1}{3}}$ in gain quantization

Taniguchi et al. [54] has provided the following probability density function of the eigenvalues of a MIMO channel,

$$f(\lambda_e) = \frac{1}{(L(e)-1)!}\frac{\lambda_e^{L(e)-1}}{\beta^{L(e)}} exp\left(-\frac{\lambda_e}{\beta}\right) \tag{B.1}$$

Here, $\lambda$ denotes an eigenvalue of the Wishart matrix (i.e., $\mathbf{H}^H\mathbf{H}$ or $\mathbf{H}\mathbf{H}^H$). $e$ denotes the order of the eigenvalue. $L(e) = (M-e)(N-e)$. $\beta$ is a constant whose value is given through the following equation,

$$\beta = \frac{\tilde{\lambda_e}}{L(e)} \tag{B.2}$$

Here $\tilde{\lambda_e}$ is the mean of the eigenvalue. (B.1) provides the probability distribution function of the eigenvalue of the Wishart matrix, $\lambda_e$. In our proposed algorithm, we are trying to quantize the singular values of the gaussian matrix, $g$. Now, $\lambda_e = g^2$.

Using Jacobian transformation [52], the probability distribution of the singular values

of the gaussian matrix can be found as follows,

$$f_g(r) = \frac{1}{(L(e)-1)!}\frac{(r^2)^{L(e)-1}}{\beta^{L(e)}} exp\left(-\frac{r^2}{\beta}\right) 2r \tag{B.3}$$

Therefore,

$$||f_g(r)||_{\frac{1}{3}} = \frac{2}{(L(e)-1)!}\frac{1}{\beta^{L(e)}}\left(\int_0^\infty r^{\frac{2L(e)-1}{3}} exp\left(-\frac{r^2}{3\beta}\right) dr\right)^3 \tag{B.4}$$

From standard mathematical tables ( [4], P - 380, eqn - 662),

$$\int_0^\infty x^n exp\left(-ax^p\right) dx = \frac{\Gamma\left(\frac{n+1}{p}\right)}{pa^{\left(\frac{n+1}{p}\right)}} \tag{B.5}$$

Comparing (B.5) with (B.4), we find, $n = \frac{2L(e)-1}{3}$, $a = \frac{1}{3\beta}$, $p = 2$. Therefore,

$$\left(\int_0^\infty r^{\frac{2L(e)-1}{3}} exp\left(-\frac{r^2}{3\beta}\right) dr\right) = \frac{\Gamma\left(\frac{\frac{2L(e)-1}{3}+1}{2}\right)}{2\left(\frac{1}{3\beta}\right)^{\frac{\frac{2L(e)-1}{3}+1}{2}}} \tag{B.6}$$

$$\left(\int_0^\infty r^{\frac{2L(e)-1}{3}} exp\left(-\frac{r^2}{3\beta}\right) dr\right) = \frac{1}{2}(3\beta)^{\frac{L(e)+1}{3}}\Gamma\left(\frac{L(e)+1}{3}\right) \tag{B.7}$$

$$\left(\int_0^\infty r^{\frac{2L(e)-1}{3}} exp\left(-\frac{r^2}{3\beta}\right) dr\right)^3 = \frac{1}{8}(3\beta)^{L(e)+1}\Gamma^3\left(\frac{L(e)+1}{3}\right) \tag{B.8}$$

So, (B.4) takes the following form,

$$||f_g(r)||_{\frac{1}{3}} = \frac{2}{(L(e)-1)!}\frac{1}{\beta^{L(e)}}\frac{1}{8}3^{L(e)+1}\beta^{L(e)+1}\Gamma^3\left(\frac{L(e)+1}{3}\right) \tag{B.9}$$

$$= \frac{3\times 3^{L(e)}\beta}{4(L(e)-1)!}\Gamma^3\left(\frac{L(e)+1}{3}\right) \tag{B.10}$$

### B.1.1 Finding mean of the eigenvalues

Taniguchi et. al. [54] provided the following analytical expression of the mean of the dominant eigenvalue,

$$\tilde{\lambda}_0 = MN\left(\frac{M+N}{MN+1}\right)^{\frac{2}{3}} \tag{B.11}$$

(B.11) holds as long as $MN \leq 250$. [54] shows estimation methods to find the mean other eigenvalues of the Wishart matrix. Using these values of the mean in (4.30), one can find the appropriate $C_g$, i.e. gain constant term.

## B.2 Shape Quantization Proof

*Proof of Lemma 2:*

Using (4.24),

$$Pr[\min_{i\in N}||\mathbf{s}-\hat{\mathbf{s}}_i||^2 \leq b] = \left(1 - \frac{(2M-1)C_{2M-1}\int_0^{cos^{-1}(1-0.5b)} sin^{2M-2}\phi d\phi}{2MC_{2M}}\right)^N \tag{B.12}$$

$$= \left(1 - K_1 \int_0^{cos^{-1}(1-0.5b)} sin^{2M-2}\phi d\phi\right)^N \tag{B.13}$$

$$\approx \left(1 - K_1 \int_0^{cos^{-1}(1-0.5b)} \phi^{2M-2} d\phi\right)^N \tag{B.14}$$

$$= \left(1 - K_2 \left(cos^{-1}\left(1-0.5b\right)\right)^{2M-1}\right)^N \tag{B.15}$$

In (B.13), we assumed $K_1 = \frac{(2M-1)C_{2M-1}}{2MC_{2M}}$. (B.14) follows from the fact that, given a large number of codevectors i.e., at high bit rate, the complementary cumulative distribution function (CCDF) is significant only for smaller values of $\phi$. For these smaller angles, we can assume $sin\phi \approx \phi$. Fig. B.1 compares the simulated shape quantization distortion with the original and approximate analytical shape quantization distortion of a $2\times 1$ $\mathbb{C}^M$ vector. The original and approximate analytical distortions were plotted using (B.13) and (B.14) respectively. The simulated distortion curve was plotted using the following way:

1. We used 1024 (i.e. $2^{10}$) unit norm random codevectors in our codebook.
2. We generated 10,000 random unit norm codevectors and assigned those to their closest codevector, stored in the codebook, in terms of Euclidean distance.
3. We found the overall Euclidean distance based quantization error.

Here, the CCDF of the original and approximate analytical expressions are superimposed with the simulated CCDF. Therefore, (B.13) and (B.14) accurately model the actual distortion. Now, for smaller distances (i.e., for small $b$), the CCDF of the original and approximate analytical expressions are very close to each other. This justifies the transition from (B.13) to (B.14). Note that, this similarity holds only for smaller

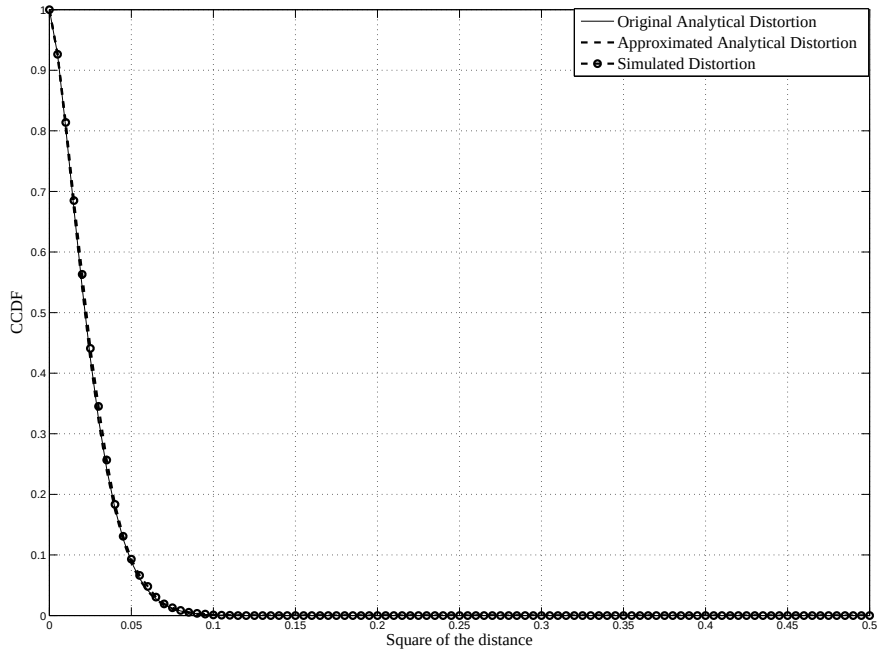



Figure B.1: Comparison of the original and approximated CCDF of the shape distortion of a 2x1 vector (10 bit quantization)

distances since $sin\phi \neq \phi$ for larger $\phi$. Therefore, although the square of the Euclidean distance of two random unit norm vectors can vary from 0 to 4, (B.15) will only hold for smaller distances. Since the CCDF of the original function is negligible outside this range, the limited boundary of (B.15) does not have any significant affect on the calculation of the expected value of the distortion. Note that, (B.15) follows from assuming $K_2 = \frac{K_1}{2M-1}$.

Now,

$$E(b) = \int_0^4 Pr[\min_{i \in N} ||\mathbf{s} - \hat{\mathbf{s}}_i||^2 \leq b] db \tag{B.16}$$

$$= \int_0^a \left(1 - K_2 \left(cos^{-1}\left(1 - 0.5b\right)\right)^{2M-1}\right)^N db \tag{B.17}$$

$$= 2\int_0^\psi \left(1 - K_2\theta^{2M-1}\right)^N sin(\theta) d\theta \tag{B.18}$$

$$\approx 2\int_0^{\psi}\left(1-K_2\theta^{2M-1}\right)^N\theta d\theta \tag{B.19}$$

$$\approx 2\int_0^{1}\left(1-K_2\theta^{2M-1}\right)^N\theta d\theta \tag{B.20}$$

$$= 2\int_0^{1}\left(\sum_{i=0}^{N}\binom{N}{i}(-1)^iK_2^i\theta^{i(2M-1)+1}\right)d\theta \tag{B.21}$$

$$= 2\sum_{i=0}^{N}\frac{\binom{N}{i}(-1)^iK_2^i}{i(2M-1)+2} \tag{B.22}$$

Although the original value of the expected distortion ranges from between 0 and 4 in (B.16), (B.17) holds only for small values $d$ due to the results explained in the previous section. The value of $d$ can be approximated as long as it does not have significant affect on the expected value. In (B.18) we assumed, $\theta = (cos^{-1}(1-0.5b))$. Therefore, $\psi = (cos^{-1}(1-0.5b))$. Since only smaller angles of $\theta$ contribute to $E(b)$, we assumed $\sin\theta \approx \theta$ in (B.19). In (B.21), we assumed $\psi = 1$ to simplify the other calculations.

Fig. B.2 justifies the approximations that we used in the derivations of shape distortion calculation. Here, approx1 and approx2 denote $sin(\theta) \approx \theta$ (ref: eq. B.19) and $\psi \approx 1$ (ref: eq. B.20) respectively. As Fig. B.2 shows, the three curves are superimposed with each other. Therefore, our justifications are valid, especially for high bit rate quantization.

Applying $\binom{N}{i} = \frac{(-1)^i(-N)_i}{i!}$, where $(-N)_i = \frac{\Gamma(-N+i)}{\Gamma(-N)}$ [3], (B.22) takes the following form,

$$\sum_{i=0}^{N}\frac{(-1)^i(-N)_i(-1)^iK_2^i}{i!(i(2M-1)+2)} = \frac{2}{2M-1}\sum_{i=0}^{N}\frac{(-N)_iK_2^i}{i!(i+\frac{2}{2M-1})} \tag{B.23}$$

$$= \frac{2}{2M-1}\frac{N!}{\frac{2}{2M-1}\left(1+\frac{2}{2M-1}\right)_N}K_2^{\frac{-2}{2M-1}} \tag{B.24}$$

$$= \frac{N!\Gamma\left(1+\frac{2}{2M-1}\right)}{\Gamma\left(N+1+\frac{2}{2M-1}\right)}K_3 \tag{B.25}$$

$$= \frac{N\Gamma(N)\Gamma\left(\frac{2M+1}{2M-1}\right)}{\Gamma\left(N+\frac{2M+1}{2M-1}\right)}K_3 \tag{B.26}$$

$$= N\beta\left(N,\frac{2M+1}{2M-1}\right)K_3 \tag{B.27}$$

(B.24) was found using ( [21], 6.6.8). In (B.25), we assumed $K_3 = K_2^{-\frac{2}{2M-1}}$. (B.27) was

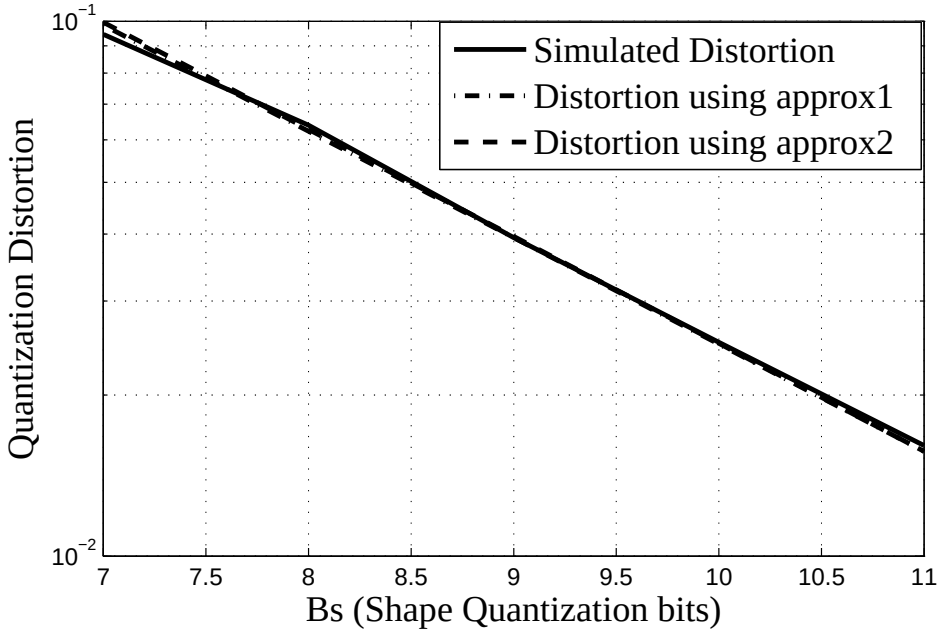



Figure B.2: Justification of the approximations used in Shape quantization distortion calculation

obtained using the relation between gamma and beta function, $\beta(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ [44]. Following a similar work of Jindal [29], we find,

$$N\beta\left(N, \frac{2M+1}{2M-1}\right) = 2^B \frac{\Gamma(2^B)\Gamma(1+\frac{2}{2M-1})}{\Gamma(2^B+1+\frac{2}{2M-1})} \tag{B.28}$$

$$\leq 2^B \frac{\Gamma(2^B)}{\Gamma(2^B+1+\frac{2}{2M-1})} \tag{B.29}$$

$$= \frac{\Gamma(2^B+1)}{\Gamma(2^B+1+\frac{2}{2M-1})} \tag{B.30}$$

The preceding inequality in (B.29) is justified with the following reasoning: due to the convexity of the gamma function [29] and the fact that $\Gamma(1) = \Gamma(2) = 1$, $\Gamma(x) \leq 1$ for $1 \leq x \leq 2$ . Let, $y = 2^B + \frac{2}{2M-1}$, $t = 1 - \frac{2}{2M-1}$, so that, $y+t = 2^B+1$, $y+1 = 2^B+1+\frac{2}{2M-1}$. By applying Kershaw's inequality for the gamma function [31],

$$\frac{\Gamma(y+t)}{\Gamma(y+1)} < \left(y+\frac{t}{2}\right)^{t-1} \quad \forall\, y > 0\,,\, 0 < t < 1 \tag{B.31}$$

Using (B.31),

$$\frac{\Gamma(2^B+1)}{\Gamma(2^B+1+\frac{2}{2M-1})} < \left(2^B+\frac{2}{2M-1}+0.5-\frac{1}{2M-1}\right)^{\frac{-2}{2M-1}} \tag{B.32}$$

$$= \left(2^B+\frac{1}{2M-1}+0.5\right)^{\frac{-2}{2M-1}} \tag{B.33}$$

$$< 2^{\frac{-2B}{2M-1}} \tag{B.34}$$

Using (B.34) and the value of $K_3$ we find,

$$2\sum_{i=0}^{N}\frac{(-1)^i(-N)_i(-1)^iK_2^i}{i!(i(2M-1)+2)} < \left(\frac{C_{2M-1}}{2MC_{2M}}\right)^{-\frac{2}{2M-1}} 2^{\frac{-2B}{2M-1}} \tag{B.35}$$

Using the values of $C_{2M-1}$ and $C_{2M}$ one can obtain,

$$E(b) \leq C_s 2^{\frac{-2B_s}{2M-1}} \tag{B.36}$$

Here, $C_s = \left(\frac{\pi^{\frac{2M-1}{2}}\Gamma(M)}{2\pi^M\Gamma\left(\frac{2M-1}{2}+1\right)}\right)^{\frac{-2}{2M-1}}$ is a constant with respect to $B_s$.

## B.3 Optimal Bit Allocation Proof

Taking the 1st and 2nd order derivatives of (4.30), we find,

$$\frac{dD}{dB_s} = \bar{C}_s(\ln 2)2^{-\frac{2B_s}{2M-1}}\left(-\frac{2}{2M-1}\right) + C_g(\ln 2)\left(2^{-2(B-B_s)}\right)2 \tag{B.37}$$

$$\frac{d^2D}{d^2B_s} = \bar{C}_s(\ln 2)^2 2^{-\frac{2B_s}{2M-1}}\left(-\frac{2}{2M-1}\right)^2 + C_g(2\ln 2)^2\left(2^{-2(B-B_s)}\right) \tag{B.38}$$

From (B.38), $\frac{d^2D}{d^2B_s} \geq 0$. Therefore, the optimal bit allocation problem is convex [7]. Now, equating the 1st derivative to be zero,

$$\frac{\bar{C}_s}{2M-1}2^{\frac{-2B_s}{2M-1}} = C_g 2^{-2(B-B_s)} \tag{B.39}$$

$$2^{-2B+2B_s+\frac{2B_s}{2M-1}} = \frac{\bar{C}_s}{C_g(2M-1)} \tag{B.40}$$

$$2B_s + \frac{2B_s}{2M-1} - 2B = \log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right) \tag{B.41}$$

$$\frac{2MB_s}{2M-1} = B + \frac{1}{2}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right) \tag{B.42}$$

$$B_s = \frac{2M-1}{2M}B + \frac{2M-1}{4M}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right) \tag{B.43}$$

Therefore, at the optimal point,

$$B_s = \frac{2M-1}{2M}B + \frac{2M-1}{4M}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right) \tag{B.44}$$

$$B_g = \frac{1}{2M}B - \frac{2M-1}{4M}\log_2\left(\frac{\bar{C}_s}{C_g(2M-1)}\right) \tag{B.45}$$

# Bibliography


[1] IEEE standard 802.16e-2005. part 16: Air interface for fixed and mobile broadband wireless access systems-ammendment for physical and medium access control layers for combined fixed and mobile operation in licensed band., 2005.

[2] J. G. Andrews, A. Ghosh, and R. Muhammad. *Fundamentals of WiMax*. Prentice Hall, NJ, 2007.

[3] C. K. Au-Yeung and D. J. Love. On the performance of random vector quantization limited feedback beamforming in a miso system. *IEEE Transactions on Wireless Communications*, 6(2):458–462, 2007.

[4] W. H. Beyer. *CRC Standard Mathematical Tables, 26th Ed.* CRC, FL, 1981.

[5] F. Boccardi, H. Huang, and M. Trivellato. Mutliuser eigenmode transmission for MIMO broadcast channels with limited feedback. In *Proc. IEEE SPAWC 2007*, June 2007.

[6] H. Boche and M. Schubert. A general duality theory for uplink and downlink beamforming. In *Proc. IEEE VTC'2002*, volume 1, pages 87–91, September 2002.

[7] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University, Cambridge, UK, 2004.

[8] C. Chae, D. Mazzarese, T. Inoue, and R. W. Heath. Coordinated beamforming for the multiuser MIMO broadcast channel with limited feedforward. *IEEE Transactions on Signal Processing*, 56:6044–6056, 2008.

[9] C. Chae, D. Mazzarese, N. Jindal, and R. W. Heath Jr. Coordinated beamforming with limited feedback in the MIMO broadcast channel. *IEEE Journal of Selected Areas in Communications*, 26:1505–1515, October 2008.

[10] J. H. Conway, R. H. Hardin, and N. J. A. Sloane. Packing lines, planes, etc. packings in Grassmanian spaces. *Experimental Mathematics*, 5:139–159, 1996.

[11] A. D. Dabbagh and D. J. Love. Multiple antenna MMSE based downlink precoding with quantized feedback or channel mismatch. *IEEE Transactions on Communications*, 7:1859–1868, November 2008.

[12] N. Dharamdial and R. S. Adve. Efficient feedback for precoder design in single- and multi-user MIMO systems. In *Conf. on Info. Sci. and Sys.*, March 2005.

[13] M. Ding. *Multiple-input multiple-output system design with imperfect channel knowledge*. PhD thesis, Queen's University, 2008.

[14] M. Ding and S. D. Blostein. Uplink-downlink duality in normalized mse or sinr under imperfect channel knowledge. In *Proc. IEEE GLOBECOM' 2007*, pages 3786–3790, December 2007.

[15] A. Gersho and R. M. Gray. *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, MA, 1991.

[16] J. D. Gibson. Adaptive prediction in speech differential encoding systems. *Proc. of the IEEE*, 68:488–525, April 1980.

[17] A. Goldsmith. *Wireless Communications*. Cambridge University Press, MA, 2005.

[18] A. Gupta and S. Nadarajah. Order statistics and records. In A. Gupta and S. Nadarajah, editors, *Handbook of Beta Distribution and Its Applications*, pages 89–96. Marcel Dekker, Inc, New York, 2004.

[19] J. C. Haartsen. Impact of non-reciprocal channel conditions in broadband TDD systems. In *Proc. IEEE PIMRC 2008*, pages 1–5, September 2008.

[20] J. Hamkins and K. Zeger. Gaussian source coding with spherical codes. *IEEE Transactions on Information Theory*, 48:2980–2989, November 2002.

[21] E. R. Hansen. *A Table of Series & Products*. Prentice Hall, Inc., NJ, 1975.

[22] S. Haykin. *Adaptive Filter Theory*. Prentice Hall, NJ, 2002.

[23] F. Hiai and D. Petz. Asymptotic freeness almost everywhere for random matrices. *Acta Sci. Math. (Szeged)*, 66:801–826, 2000.

[24] K. Huang, R. W. Heath, and J. G. Andrews. Limited feedback beamforming over temporally-correlated channels. *IEEE Trans. on Sig. Proc.*, 57(5):1959–1975, 2009.

[25] M. N. Islam and R. S. Adve. Linear transceiver design in a multiuser MIMO system with quantized channel state information. In *Proc. IEEE ICASSP 2010*, pages 3410–3413, March 2010.

[26] M. N. Islam and R. S. Adve. SMSE precoder design in a multiuser miso system with limited feedback. In *Proc. Queen's Biennial Symposium on Communications 2010*, pages 352–356, May 2010.

[27] M. N. Islam and R. S. Adve. Transceiver design using linear precoding in a multiuser system with limited feedback. IET Journals on Communications, accepted. arXiv:1001.2307v1 [cs.IT].

[28] N. S. Jayant. Adaptive delta modulation with a one-bit memory. Technical report, Bell System Technical Journal, NJ, March 1970.

[29] N. Jindal. MIMO broadcast channels with finite-rate feedback. *IEEE Transactions on Communications*, 52:5045–5060, November 2006.

[30] N. Jindal. Antenna combining for the MIMO downlink channel. *IEEE Transactions on Wireless Communications*, 7:3834–3844, October 2008.

[31] D. Kershaw. Some extensions of the W. Gautschi's inequalities for the gamma function. *Mathematics of Computation*, 41(164):607–611, October 1983.

[32] A. M. Khachan, A. J. Tenenbaum, and R. S. Adve. Linear processing for the downlink in multiuser MIMO systems with multiple data streams. In *Proc. IEEE ICC'2006*, volume 9, pages 4113–4118, June 2006.

[33] B. Khoshnevis and W. Yu. Limited feedback multi-antenna quantization codebook design part I: Single-user channels. IEEE Transactions on Signal Processing, submitted for publication. arXiv:1003.2257v1 [cs.IT].

[34] B. Khoshnevis and W. Yu. Limited feedback multi-antenna quantization codebook design part II: Multiuser channels. IEEE Transactions on Signal Processing, submitted for publication. arXiv:1003.2259v1 [cs.IT].

[35] K. Kim, H. Kim, and D. J. Love. Utilizing temporal correlation in multiuser MIMO channels. In *Proc. IEEE ASILOMAR 2008*, November 2008.

[36] T. Kim, D. J. Love, B. Clerckx, and S. J. Kim. Differential rotation feedback MIMO system for temporally correlated channels. In *Proc. IEEE GLOBECOM 2008*, December 2008.

[37] M. Kobayashi, G. Caire, and N. Jindal. How much training and feedback are needed in MIMO broadcast channels? In *Proc. IEEE ISIT 2008*, pages 2663–2667, July 2008.

[38] Y. Linde, A. Buzo, and R. M. Gray. An algorithm for vector quantizer design. *IEEE Trans. on Comm.*, 28:84–95, January 1980.

[39] S. P. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28:129–137, March 1982.

[40] D. J. Love and R. W. Heath. Feedback techniques for MIMO channels. In G. Tsoulos, editor, *MIMO System Technology for Wireless Communications*. Taylor and Francis Group, Boca Raton, FL, 2006.

[41] D. J. Love, R. W. Heath Jr., and T. Strohmer. Grassmanian beam-forming for multiple-input multiple-output wireless systems. *IEEE Transactions on Information Theory*, 49:2735–2747, October 2003.

[42] Mathworks. The mathworks website. http://www.mathworks.com/access/helpdesk/help/toolbox/stats/kmeans.html, 2010.

[43] J. Max. Quantizing for minimum distortion. *IRE Transactions on Information Theory*, 6:7–12, 1960.

[44] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Process*. McGraw-Hill Companies, Inc., NY, 2002.

[45] F. Hiai & D. Petz. *The semicircle law, free random variables, and entropy*. American Mathematical Society, RI, 2000.

[46] N. Ravindran and N. Jindal. MIMO broadcast channels with block diagonalization and finite rate feedback. In *Proc. IEEE ICASSP 2007*, April 2007.

[47] J. C. Roh and B. D. Rao. An efficient feedback method for MIMO systems with slowly time-varying channels. In *Proc. IEEE WCNC 2004*, March 2004.

[48] J. C. Roh and B. D. Rao. Transmit beamforming in multiple-antenna system with finite rate feedback: A VQ-based approach. *IEEE Transactions on Information Theory*, 52:1101–1112, March 2006.

[49] M. Schubert, S. Shi, E. A. Jorswieck, and H. Boche. Downlink sum-MSE transceiver optimization for linear multi-user mimo systems. In *Proc. Asilomar Conf. on Signals, Systems and Computers*, volume 9, page 14241428, September 2005.

[50] M. B. Shenouda and T. N. Davidson. On the design of linear transceivers for multiuser systems with channel uncertainty. *IEEE Journal of Selected Areas in Communications*, 26:1015–1024, August 2008.

[51] S. Shi and M. Schubert. MMSE transmit optimization for multi-user multi-antenna systems. In *Proc. IEEE ICASSP'2005*, volume 3, pages 409–412, March 2008.

[52] G. Strang. *Linear Algebra & Its Applications*. Harcourt Brace Jovanovich Publishers, CA, 1988.

[53] R. Stroh. *Optimum and Adaptive Differential Pulse Coded Modulation*. PhD thesis, Polytechnic University, 1970.

[54] T. Taniguchi, S. Sha, and Y. Karasawa. Analysis and approximation of statistical distribution of eigenvalues in i.i.d. MIMO channels under Rayleigh fading. *IEICE Trans. on Info. Th. and Its Appl.*, E91-A:2808–2817, October 2008.

[55] A. J. Tenenbaum and R. S. Adve. Joint multiuser transmit-receive optimization using linear processing. In *Proc. IEEE ICC'2004*, volume 1, pages 588–592, 2004.

[56] A. J. Tenenbaum, R. S. Adve, and Y. Yuk. Channel prediction and feedback in multiuser broadcast channels. In *Proc. CWIT*, 2009.

[57] A. J. Tenenbaum and Raviraj S. Adve. Minimizing sum-MSE implies identical downlink and dual uplink power allocations, submitted for publication. arXiv:0912.3323v1 [cs.IT].

[58] M. Trivellato, H. Huang, and F. Boccardi. Antenna combining and codebook design for MIMO broadcast channel with limited feedback. In *Proc. Asilomar Conf. on Sig. and Systems*, pages 302–308, November 2007.

[59] Y. R. Zheng and C. Xiao. Simulation models with correct statistical properties for rayleigh fading channels. *IEEE Transactions on Communications*, 51(6):920–928, 2003.